

Identification of Peer Effects with Missing Peer Data: Evidence from Project STAR

Aaron Sojourner*

November 6, 2010

Abstract

This paper studies peer effects on student achievement among first graders randomly assigned to classrooms in Tennessee's Project STAR. The analysis uses previously unexploited pre-assignment achievement measures available for 60 percent of students. Data are not missing at random, making identification challenging. The paper develops new ways, given random assignment of individuals to classes, to identify peer effects without imposing other missing-data assumptions. Estimates suggest positive effects of mean peer lagged achievement on average. Allowing heterogeneous effects, evidence suggests lower-achieving students benefit more than higher-achieving students do from increases in peer mean. Further, the bias in a widely used, poorly understood peer-effects estimator is analyzed, implying that caution is warranted in interpreting many peer-effects estimates extant in the literature.

JEL Classifications: C2, I21, J13

*University of Minnesota, Carlson School of Management, asojourn@umn.edu. This paper benefited from many helpful insights from Charles Manski, Christopher Taber and Greg Duncan. Thanks also to Charles Achilles, Jayne Boyd-Zaharias, Matthew Genztkow, Larry Hedges, Kyoo Il Kim, Colleen Manchester, Jason O'Connor, Steve Pischke, Christopher Rhoads, James Roberts, Elizabeth Tipton, and seminar participants at many venues for generous assistance at various stages. All errors are mine. This work was supported by a fellowship from the U.S. Department of Education Institute for Education Science.

1 Introduction

Imagine yourself in the position of an elementary school principal about to start a new school year. You can access records of each student’s background characteristics and past years’ test scores. You must decide how to divide the students into classes. If each student’s classmates affect her outcomes, then your class composition decision has consequences both for students and for society. Answers to some empirical questions would help with your decision. How can students’ outcomes be expected to differ depending on their peers? What trade-offs do you face in choosing one class composition policy versus another? Although not sufficient for choosing optimal policy, understanding how peer effects work would be useful. A large literature on peer effects in education aims to generate insight into how classroom peers affect student outcomes.

In application, this paper examines peer effects on academic achievement among first graders randomly assigned to their classrooms and teachers as part of Tennessee’s Student/Teacher Achievement Ratio Project (Project STAR), one of America’s largest education experiments ever conducted. The analysis draws on previously unexploited measures of kindergarten achievement taken before random assignment to first grade classes and available for about 60 percent of this sample. Data are not missing at random. This paper studies the effects of peer lagged achievement on own first grade achievement.

Peer effects also play a prominent role in research on workplace teams, health habits, marketing, neighborhood influences, price bubbles, crime, microfinance, networks, and elsewhere. This paper contributes methodologically to the larger peer-effects literature by developing a new way to exploit random assignment of individuals to groups to overcome an otherwise severe problem caused by missing data on the causally relevant peer variables.

Peer-effects researchers confront at least two methodological challenges: how to deal with possible selection into peer groups and how to deal with missing data. Although it is well known that random assignment of individuals to peer groups helps avoid selection problems, this paper points out how it can help overcome missing data problems as well.

Peer-effects researchers must separate the causal effects of peers from “correlated effects,”

the influence exerted by omitted variables that are correlated with peer measures [Manski (1993)]. Researchers analyzing normal administrative data sets from school districts do not generally understand the process of student and teacher sorting well enough to model selection credibly. Consequently, estimates can suffer from omitted-variable bias, mistaking the influence of unobserved factors for the causal influence of peers. Though difficult to address, this challenge is well understood and receives a great deal of attention in empirical papers. In this paper’s empirical analysis, random assignment of students and teachers to classes helps meet this challenge.

Researchers must also deal with the challenge posed by any missing data on variables theorized to influence peers. Methods and intuitions for dealing with missing data that were developed in other contexts do not translate immediately to peer-effects research. Peer-effects research differs because missing information about one individual does not remain isolated to that individual’s observation. Rather, it spills over as missing peer information for all of the individual’s peers as well. The missing-data problem is particularly severe in the data studied here, as about 40 percent of students are missing lagged achievement scores and students select into missingness.

The challenge posed by missing data on peer-influencing covariates rarely receives attention in the theoretical or empirical literature, though this kind of missing data pervades empirical work. Many studies deal with missing data using an estimation procedure that this paper terms an *individual-deletion procedure* (IDP): delete individuals with missing covariate data and carry out the analysis as if the remaining individuals represented the population. Consider these examples from recent studies of peer effects in education. In their study of K-12 schools, Vigdor and Nechyba (2007) drop about 20 percent of students in this way. Lefgren (2004) drops between 8 and 10 percent of various cohorts. Atkinson et al. (2008), Burke and Sass (2008), and Zabel (2008) do the same without specifying the percentage of individuals dropped. Foster (2006) studies peer effects at the University of Maryland. She drops at least 10 percent of individuals due to missing data. Arcidiacono and Nicholson (2005) drop about 35 percent from their study of peer effects in medical schools. Though convenient, the prop-

erties of IDP estimates are not well understood. IDP estimators are often employed without much comment or theoretical justification. The current paper shows that IDP estimators are generally biased and inconsistent and relates this bias to the interplay between the data missingness and group formation processes.

This paper develops a new way to handle missing peer data that delivers unbiased estimates of peer effects under certain conditions. The paper shows that peer effects are point identified in the most commonly used model, the linear-in-peer-means model, when peer groups are formed by random assignment or under any other conditions that generate mean independence for the expectation of the mean value of missing-data peers' causal variable. The result holds without any other conditions on the distribution of missing data. Missing data are not restricted to be missing completely at random (MCAR) or missing at random conditional on other observable variables (MAR). This facilitates credible inference when individuals select into missingness. The result comes from decomposing causal peer variables into observed and unobserved portions. This analysis suggests a simple procedure for obtaining unbiased estimates of effects of some properties of the distribution of peer characteristics. A set of closely related estimators is also proposed, which have substantially less variance at the expense of potentially introducing some bias. Analytic and Monte Carlo results are used to describe this trade-off. Empirical analysis is done with a range of estimators in this set.

Further, this paper shows that peer effects are partially identified when peer groups are formed by an unknown assignment process. The paper presents estimates of an inner bound on the range of peer effects consistent with the observed data and the model using a method that hinges on considering all feasible distributions for the missing data. Until now, this type of partial identification analysis has not been adapted to the peer effects context.

This paper's empirical contribution derives from overcoming both methodological challenges, thereby, generating credible insight into the operation of peer effects in first grade classrooms. Drawing on the identification results, the effect of mean peer lagged achievement for all students is estimated. Estimates assuming homogeneous effects suggest that, on average, end-of-first-grade achievement rises moderately as mean peer lagged achievement

risers. The paper then permits heterogeneous peer effects, allowing the strength of peer effects to differ between students with low, middle, and high levels of lagged achievement. The evidence weakly suggests that lower-achieving students benefit more than higher-achieving students from increases in the peer mean. When we entertain the possibility that students may be assigned to classes nonrandomly, we see that missing peer data make the observed data consistent with a wide range of peer effects.

Concurrent work by Ammermueller and Pischke (AP) is the closest paper in the literature, and we have all benefited from our dialogue. They also consider the problem of inference on peer effects with missing data on the causally relevant peer variable. By a different mode of analysis and using narrower conditions, they propose a correction to the IDP estimator. Early versions of their paper relied on MCAR assumptions [Ammermueller and Pischke (2008)]; based partly on insights from this paper, later versions relax MCAR and rely on random assignment alone [Ammermueller and Pischke (2009)]. Their estimator is biased and inconsistent across multiple schools, although the bias is small under certain conditions. The performance of the IDP and the AP-corrected IDP estimators are compared to those developed here analytically and through simulations. They also consider issues not considered here, such as measurement error in the peer-influencing variable. In application, our papers focus on different populations and variables.

The remainder of the paper is organized as follows. Section 2 generalizes the linear-in-peer-means model to allow for missing data on covariates that influence peers' outcomes. Section 3 studies identification of the model, develops new methods, and describes properties of peer-effects estimators under various conditions on the peer group formation and data missingness processes. Section 4 describes the STAR data, explains why data are not credibly missing at random, and provides evidence on random assignment of students and teachers to classes. Section 5 describes empirical results. Section 6 highlights the paper's limitations, describes the next steps in this research program, and concludes. Connections to the literature are discussed throughout.

2 Model

A standard linear-in-peer-means model of achievement production is modified to study what happens when data on the individual covariates that influence peers' outcomes are missing. This section defines the population, production function, and variables.

Population and sample. Index individual students by $i = 1, 2 \dots I$. In any given academic year, each student is assigned to a class in a school. Classes are indexed by $c = 1, 2 \dots C$ and the number of students in class- c is N_c . Schools are indexed by $s = 1, 2 \dots S$, and each contains multiple classes.¹

Production. Each individual- i 's outcome (y) is produced according to this function of mean leave-me-out classroom-peer characteristics (\bar{v}), own and class wide characteristics (x), a school fixed effect, and within-school unobserved influences (e):

$$y = \beta \bar{v} + f(x) + \gamma_s + e \quad (1)$$

Starting with a nonparametric function of own and class wide characteristics f allows one to clarify where identification requires linearity ($f(x) \equiv \pi x$) and where it does not. Estimation will always assume linear production. Most empirical papers on peer effects use a version of this production function. It identifies the combination of endogenous and exogenous effects as β without attempting to separate them [Manski (1993)].

Variables. The outcome, $y \in Y \subseteq \mathbf{R}$, is continuous, and its value for individual- i in classroom- c in school- s is y_{sci} . In this application, y is the end-of-first-grade academic achievement percentile.

Each student has a vector of covariates $x \in X \subseteq \mathbf{R}^K$ that influence her own outcomes. This includes student characteristics fixed at the time of peer group formation, such as birth date or pre-assignment measures of achievement. It may also include characteristics of the classroom, such as the teacher's years of experience or class size.

To capture peer effects, each individual's y also depends on her peers' x . Consider some

¹To apply the model to other contexts, consider individuals in peer groups (classes) drawn from multiple, separate populations (schools).

known function $g : X \rightarrow V \subseteq \mathbf{R}^M$ for all $x \in X$ and define each student's peer-influencing variable $v \equiv g(x)$, which could be a vector. For any student- i in class- c , peers affect her y_{sci} via her leave-me-out classroom mean \bar{v} ,

$$\bar{v}_{sci} \equiv (N_c - 1)^{-1} \sum_{j \neq i, j \in c} g(x_{scj}). \quad (2)$$

Substantively, this restriction permits analysis of just those properties of the distribution of peer x that can be expressed as additively separable functions of individual peers' x and is necessary for the decomposition used to break \bar{v}_{sci} into observed and missing portions.

Define $z \in \{0, 1\}$ to indicate the observability of an individual's $v = g(x)$. If $z_{sci} = 0$, individual- i 's v_{sci} is not observable due to missing data on components of x_{sci} that matter for i 's peers. Otherwise, if the missing data were on components of x that did not affect peers, there would be no missing peer data problem. If $z_{sci} = 1$, all components of x_{sci} are observed, which is sufficient for observation of v_{sci} .

In any class- c with N_c total students, let the number with fully observed data $z = 1$ be n_c . The number with missing data is $N_c - n_c$. For any student- i in class- c , some peers' v are observed and others' v are missing. The researcher knows the fraction of i 's peers with observed v :

$$p_{sci} \equiv (N_c - 1)^{-1} \sum_{j \in c, j \neq i} z_{scj}.$$

Since N_c and n_c vary across classes within school, so does p . Within class, p varies slightly depending on one's own z .

Define observed v as $v^{ob} \equiv vz$ and missing v as $v^m \equiv v(1 - z)$. The mean v among student- i 's classroom- c peers with observed v is

$$\bar{v}_{sci}^{ob} \equiv \begin{cases} [(N_c - 1)p_{sci}]^{-1} \sum_{j \in c, j \neq i} v_{scj}^{ob}, & \text{if } p_{sci} > 0 \\ 0, & \text{if } p_{sci} = 0. \end{cases}$$

The mean v among the student's peers with missing v is

$$\bar{v}_{sci}^m \equiv \begin{cases} [(N_c - 1)(1 - p_{sci})]^{-1} \sum_{j \in c, j \neq i} v_{scj}^m, & \text{if } p_{sci} < 1 \\ 0, & \text{if } p_{sci} = 1. \end{cases}$$

By definition, \bar{v}_{sci}^{ob} is always observed and \bar{v}_{sci}^m is never observed except trivially when $p_{sci} = 1$. The causal peer variable \bar{v}_{sci} can be decomposed into observed and unobserved portions using the identity

$$\bar{v}_{sci} = \bar{v}_{sci}^{ob} p_{sci} + \bar{v}_{sci}^m (1 - p_{sci}). \quad (3)$$

This decomposition will be useful in dealing with the missing peer data. It also highlights the types of functions that are permissible for g and the properties of the distribution of peer x that are amenable to study in this framework.²

The model is more flexible than it may first appear and includes many commonly studied aspects of peers. Consider three examples of using this framework to study the effect of different aspects of peers. Let the first component of x_{sci} ($x_{1,sci}$) be own lagged y_{sci} . First, to study the effect of mean peer lagged achievement, let $g(x_{sci}) \equiv x_{1,sci}$. Second, to study the effect of the fractions of peers with lagged achievement scores in the top and bottom terciles, with middle tercile as the omitted category, let $T1 \equiv [1, 33.3]$, $T3 \equiv [66.7, 99]$ and $g(x_{sci}) \equiv (1(x_{1,sci} \in T1), 1(x_{1,sci} \in T3))'$. Third, to study the effect of the fraction of peers female, let g return a female indicator. Polynomials, logs, or other functions of x_{sci} are also allowed. In each case, equation (2) governs and the causal peer variable \bar{v}_{sci} is additively separable in i 's peers' x_{scj} . Missing peer data pose an open and difficult challenge in analyzing the effects of nonseparable properties of the peer distribution.³

²To verify your understanding of the notation, suppose $N_c = 3$ for some class- c and suppose v is observable for students 1 and 2 ($z_{sc1} = 1, z_{sc2} = 1$) but not for student 3 ($z_{sc3} = 0$). Then, $p_{sc1} = p_{sc2} = \frac{1}{2}$ and $p_{sc3} = 1$. As described by equation (3), mean peer influence for student 1 is $\bar{v}_{sc1} = \frac{1}{2}(\bar{v}_{sc1}^{ob} + \bar{v}_{sc1}^m) = \frac{1}{2}(v_{sc2} + v_{sc3})$.

³For instance, suppose each student's outcome depends on minimum peer lagged achievement due to disruptiveness in behavior or curriculum. If, in each class, we are missing v for one student, information about all the other students' v does little to pin down the minimum. Formally, this effect of the peer minimum would be operationalized as $v_{scj} = \min_{k \neq i, k \in c} x_{1,scj}$. This violates the separability condition since $\forall j, k \in c, \forall j, k \neq i, v_{scj}$ depends on x_{scj} as well as on x_{scj} . The standard deviation of $x_{1,scj}$ is also ruled out since it requires computing the squared deviation of each individual $x_{1,scj}$ from the mean of all $x_{1,scj}, \forall j, k \neq i, j, k \in c$. This mean is not known if $p_{sci} < 1$. Some previous studies of the effect of the standard deviation of peer lagged achievement use IDP to deal with missing data and so deliver biased estimates [Hanushek et al. (2003), Vigdor and Nechyba (2007), Zabel (2008)]. Neither percentiles nor order statistics of the distribution of peer x are permissible in the current framework but remain open for future research. Understanding the effect of the

The school fixed effect γ_s is defined as the average, additive unobserved influence among students in each school. Conceptually, γ_s is generated by two processes, which we do not separate. First, a selection process matches students to schools at the beginning of the year. Students are not assigned to schools randomly, so this generates unobserved differences between school populations. Figure 1 gives a sense of how different the Project STAR schools are from one another in terms of average returning student achievement and fraction of students returning. Second, unobserved school quality differences operate during the year. Non-shared, within-school unobserved influences are captured in $e \in \mathbf{R}$ with $E[e|s]$ normalized to 0.

Define $\theta \equiv (\beta, f, \gamma)$ to represent elements of the production function. Neither θ nor the conditional distribution of individual unobserved influences $P(e|\bar{v}^{ob}, \bar{v}^m, x, s, z, p)$ is known.

If one assumes that the unobserved influences e and the missing x values have known parametric distributions and that the x data are missing at random conditional on observable data, one can achieve point identification by integrating out the missing data. Such a model can be estimated by maximum likelihood methods. However, the focus here is on obtaining identification absent these strong assumptions, which have questionable credibility in this application and many others.

spread of peer achievement would be particularly interesting because, for the school principal, moving from a policy of achievement mixing to one of achievement tracking would increase the spread of each student's peers' lagged achievement in addition to introducing changes in the peer mean.

3 Identification

This section explores identification of θ under various conditions on the peer group formation and data missingness processes. Throughout, we suppose the researcher has an interest in $P(y|\bar{v}, x, s)$, which is equivalent to knowledge of θ . It describes the distribution of outcomes among students with individual, observable pre-assignment characteristics x who are assigned to a class with peers of type \bar{v} . This knowledge in the presence of peer effects could help a school principal raise welfare through classroom composition policy. If the principal can take the distribution of students as given and if θ captures a stable structural relationship between peers and outcomes, this is all the principal needs. In more general, more realistic contexts, families and schools can respond to one another's decisions [Epple and Romano (1998), Epple et al. (2002), Hidalgo-Hidalgo (2007), Carrell et al. (2010)]. In that case, knowledge of θ is useful but not sufficient for optimal policy.

3.1 Complete data, $Pr(z = 1) = 1$

As a prelude to the problem of missing peer data, this section discusses the most common way researchers identify peer effects given complete data. The primary concern is potential selection into peer groups. A mean independence condition identifies θ if all data are observed:

$$MI : E[e|\bar{v}, x, s] = E[e|s].$$

Here, the researcher observes $P(y, \bar{v}^{ob}, x, s|z = 1, p = 1) = P(y, \bar{v}, x, s)$. This implies $P(y|\bar{v}, x, s)$, which identifies θ under MI . Mean regression of y on (\bar{v}^{ob}, x, s) identifies β , f , and γ_s :

$$\begin{aligned} E[y|\bar{v}^{ob}, x, s, z = 1] &= E[y|\bar{v}, x, s] \\ &= \beta\bar{v} + f(x) + \gamma_s + E[e|\bar{v}, x, s] \\ &= \beta\bar{v} + f(x) + \gamma_s. \end{aligned}$$

The credibility of inference hinges on the credibility of MI . It overcomes the first methodological challenge, allowing researchers to separate the causal peer effects of \bar{v} from the effects of unobserved influences e .

Consider an example when MI would not be satisfied. Suppose unobserved teacher quality were systematically matched to student achievement and, hence, to peer achievement. This might happen due to unmodeled parent, teacher, or principal influence over the student-teacher matching process. Identification of peer effects will fail, as e would not be mean independent of \bar{v} . Estimates of β would suffer from omitted-variable bias. If better teachers go to classes with higher (lower) levels of lagged student achievement, the effect of mean peer lagged achievement will be over- (under) stated.

Random assignment of students and teachers to classes has great value because it gives MI credibility. To see this, decompose within-school unobserved influences into a shared, within-class component and an idiosyncratic component, $e \equiv \kappa_c + \epsilon$. Unobserved differences in teacher quality or other classroom resources that affect everyone in class equally are captured by κ_c . Under random assignment of teachers, expected teacher quality does not depend on own and peer x , $E[\kappa_c|s] = E[\kappa_c|\bar{v}, x, s]$. Under random assignment of students, $E[\epsilon|\bar{v}, x, s] = E[\epsilon|x, s]$ since peer x is randomly assigned and independent of own ϵ . Given $E[\epsilon|x, s] = E[\epsilon|s]$, we obtain MI .⁴

In this way, Sacerdote (2001) and Zimmerman (2003) study the effect that randomly-assigned college roommates have on one another. Others have exploited the random assignment of students and teachers to classes in Project STAR to understand the effect of pre-assignment observables. Dee (2004) studies the effect of student-teacher racial matches, Whitmore (2005) the effect of fraction of peers female, and Cascio and Schanzenbach (2008) the effect of peers' ages. Unlike the present study, previous papers have not included pre-assignment achievement in own or peer x because these measures are not generally available in the STAR data.

Boozer and Cacciola (2001) and Graham (2008) also study peer effects using the STAR

⁴Formal arguments are provided in Section A of the Web Appendix available at https://netfiles.umn.edu/users/asojourn/Sojourner_Peer_WebApp.pdf.

data, but their work differs from the approach described thus far. One fundamental difference is that they study the effect of peers' y on own y rather than the effect of peers' x on own y . This is possible without pre-assignment measures of achievement but is not as useful for the school principal's problem, since y is not observable at the time classroom allocations must be decided. Also, methodologically Graham does not use mean independence conditions but develops an innovative analysis based on conditional variance restrictions.

Without random assignment of students and teachers to classes, researchers justify *MI* in other ways. They may assume that they are conditioning on sufficient observables so as to render unobservables mean independent [Ding and Lehrer (2007)]. Given multiple matches of students and teachers, student and teacher fixed effects can be used to condition out the constant, additive unobserved influences attributable to each [Betts and Zau (2004), Arcidiacono et al. (2007), Burke and Sass (2008)]. However, the credibility of this approach is undermined if schools sort students on the basis of annual fluctuations in unobservables. Rothstein (2010) finds evidence of precisely this in elementary schools. Other researchers appeal to known institutional features of the data-generating process to argue that variation in peer assignment is mean independent of unobservables [Hoxby (2000), Lavy et al. (2007), Vigdor and Nechyba (2007)]. Some papers combine both fixed effects and knowledge of the data-generating process [Hanushek et al. (2003), Cooley (2006), Hoxby and Weingarth (2007), Zabel (2008)]. All of these aim to make *MI* credible without groups being formed by random assignment of students and teachers.

3.2 New approaches given missing data

Consider the common case where each class contains individuals with missing data on the peer-influencing covariate. The researcher observes $P(y, \bar{v}^{ob}, x, s, z, p)$ across individuals. Using

equations (1) and (3), the mean regression of y on $(v^{ob}, p, x, s, z = 1)$ gives

$$\begin{aligned}
E[y|\bar{v}^{ob}, p, x, s, z = 1] &= E[\beta[\bar{v}^{ob}p + \bar{v}^m(1 - p)] + f(x) + \gamma_s + e|\bar{v}^{ob}, p, x, s, z = 1] \\
&= \beta\bar{v}^{ob}p + \beta(1 - p)E[\bar{v}^m|\bar{v}^{ob}, p, x, s, z = 1] \\
&\quad + f(x) + \gamma_s + E[e|\bar{v}^{ob}, p, x, s, z = 1].
\end{aligned} \tag{4}$$

This regression focuses on the subsample of students whose own v is observed ($z = 1$). By assuming z does not enter the production function, we imposed that the $z = 1$ subpopulation and the whole population share the same β and f .⁵ However, the distributions of x and e can differ depending on z . Without further conditions, θ is not identified. We want conditions that overcome two main challenges: selection (the distribution of e depends on (\bar{v}^{ob}, p, x)) and missing data (the distribution of \bar{v}^m depends on (\bar{v}^{ob}, p, x)). The following sections study such conditions.

3.2.1 Mean independence and connection to random assignment of students

This section describes a new way to handle missing peer data relying on two mean independence conditions, *MI1* and *MI2*, to point-identify β . Random assignment of students (both those with observed and those with missing v) and teachers to classes helps make these conditions credible, but random assignment is not necessary for the conditions to hold. Beyond this, no other conditions are imposed on the missing data, such as missing completely at random or missing at random conditional on observables. The missing and observed data could come from arbitrarily different distributions, making the analysis consistent with data missing not at random (MNAR) or nonignorability. For example, it could be that students with observed v data are higher achieving than those with missing data, so that $E[v^{ob}|s] > E[v^m|s]$. Section A of the Web Appendix formally connects the mean independence conditions to random assignment.

⁵Alternatively, this analysis could be interpreted as identifying the effect of observed peers, which may differ from the unidentified effect of missing-data peers. That is, in the last line of equation (4), we could distinguish β^{ob} as the parameter multiplying $\bar{v}^{ob}p$ and β^m as the parameter on $(1 - p)E[\bar{v}^m|\bar{v}^{ob}, p, x, s, z = 1]$. In the following analysis, β^{ob} would be identified but not β^m .

The first mean independence condition requires that any student in school- s would expect the mean of her peers' missing v (\bar{v}^m) to be the same as the average missing v in the school regardless of the realized mean v of the student's own observed-data peers (\bar{v}^{ob}), the fraction of her peers with observed v (p), her own characteristics (x), or the observability of her own v (z):

$$MI1 : E[\bar{v}^m | \bar{v}^{ob}, p, x, s, z] = E[v^m | s].$$

Among students with missing v in the school population, there is an average value for v , $E[v^m | s]$. Random assignment of students to classes within school implies the credibility of $MI1$. By random assignment of students to classes, each student in the school should expect her missing peers' mean v to equal it. Conditional on school, (\bar{v}^{ob}, p, x, z) does not affect this expectation. Although random assignment of all individuals to groups is sufficient for $MI1$, it is not necessary. $MI1$ can arise under other assignment processes. For instance, $MI1$ would also come from a deterministic process of assigning $z = 0$ individuals to classes in a way that is independent of the assignment of $z = 1$ individuals ($\bar{v}^{ob}, x, z = 1$) and the fraction of $z = 1$ individuals per class, which determines p .⁶

We also use a mean independence condition on the within-school unobserved influences of those with observed v :

$$MI2 : E[e | \bar{v}^{ob}, p, x, s, z = 1] = E[e | s, z = 1].$$

Conditional on school and own observability ($s, z = 1$), the expected unobserved influence on outcomes is the same for all students regardless of (\bar{v}^{ob}, p, x) . Random assignment of teachers gives credibility to $\kappa_c \perp (\bar{v}^{ob}, p, x)$. Conditional on $(s, z = 1)$, systematic matching of κ_c

⁶The following three examples are constructed simply to illustrate that $MI1$ can hold without random assignment of all individuals to groups. They do not describe the STAR data. First, suppose $z = 1$ individuals are randomly assigned to classes. Suppose further that all $z = 0$ individuals share a single value of $x = x^m$. In that case, $\text{Var}(x^m) = 0$, so \bar{v}^m is constant and $MI1$ holds. Second, suppose there are two values of x^m : one higher and one lower, $x^{mh} > x^{ml}$. Suppose a principal knows each $z = 0$ student's value of x^m although the econometrician does not. If the principal assigns one x^{mh} and one x^{ml} individual to each class in the school, then $MI1$ will hold. Third, suppose that one of each type is assigned to half the classes and two of each are assigned to the other half and that the choice between these two treatments is random; then $MI1$ holds. In each case, \bar{v}^m is the same for every student in the school and $MI1$ holds. More cases are possible.

to students depending on (v, x, p) would violate *MI2*. Systematic matching would create a classic omitted-variable problem. Random assignment of teachers to classes within school alleviates concern about this problem. The same arguments apply to any other influence common within class but differing between classes within schools, such as teachers' aides or computers. If they are unobserved and mean dependent on observables, they threaten identification. If they are either observed or mean independent of observables, they do not. Random assignment of students to classes ensures that students' idiosyncratic, unobserved influences ϵ are independent of (\bar{v}^{ob}, p) . Given these and a standard mean independence assumption ($E[\epsilon|x, s, z = 1] = E[\epsilon|s, z = 1]$), *MI2* follows.

MI1 – *MI2* are sufficient for point identification of β . They imply that equation (4) becomes

$$\begin{aligned}
E[y|\bar{v}^{ob}, p, x, s, z = 1] &= \beta \bar{v}^{ob} p + \beta(1 - p)E[\bar{v}^m|\bar{v}^{ob}, p, x, s, z = 1] \\
&\quad + f(x) + \gamma_s + E[e|\bar{v}^{ob}, p, x, s, z = 1] \\
&= \beta \bar{v}^{ob} p - \beta E[v^m|s]p + \beta E[v^m|s] + f(x) + \gamma_s + E[e|s, z = 1].
\end{aligned} \tag{5}$$

The production function (1) and decomposition of \bar{v} (3) give the first equality. *MI1* – *MI2* give the second.

Equation (5) is the basic expression used for identification. Consider the mean regression of y on $(\bar{v}^{ob}p, p \cdot 1_s, x, 1_s)$. Variation in $\bar{v}^{ob}p$ identifies the peer mean effect β . Variation in p within school permits recovery of each school's mean missing v , $E[v^m|s]$, from the second term and knowledge of β . It is the coefficient on $p \cdot 1_s$ divided by β . This permits identification of $E[v^m|s]$ as $n_s \rightarrow \infty$. The coefficient on each school- s dummy identifies $\beta E[v^m|s] + \gamma_s + E[e|s, z = 1]$, from which we can back out $\gamma_s + E[e|s, z = 1] \equiv \gamma_{1s}$. Using just the $z = 1$ subsample yields school fixed effects normalized to their mean unobserved influence. Estimates based on equation (5) will be referred to as *p-weight* estimators.

Point identification of θ is obtained without any conditions on the missing data beyond

MI1. Conditions such as data MCAR or MAR were not necessary. However, random assignment of $z = 0$ students and the additive separability of \bar{v} in individual peer x were essential. Identification in more general production models is possible but challenging.⁷

This analysis suggests that mean regression of y_{sci} on $(\bar{v}_{sci}^{ob}p_{sci}, x_{sci}, p_{sci}1_s, 1_s)$ conditional on $z_{sci} = 1$ will produce an unbiased estimate of β and other production parameters. However, in estimation, we are concerned with variance as well as bias. Unfortunately, inclusion of the $p_{sci}1_s$ interaction terms induces a great deal of collinearity with other regressors $(\bar{v}_{sci}^{ob}p_{sci}, 1_s)$ and requires estimation of S parameters, one for each school, in addition to the production parameters (β, f, γ) .

As interest focuses on β and we are willing to forego estimation of the S separate $E[v^m|s]$ values, we propose a set of restricted estimators, closely related to the unrestricted p -weight estimator. These greatly improve precision at the cost of introducing some bias. Below, we analytically study the bias and propose ways to deal with it. Rather than including S terms interacting p with school fixed effects, consider a model with K such interaction terms for $1 \leq K < S$. In such a model, the S schools are partitioned into K groups, a set of K dummies are created 1_k , and each student's p is interacted with her group dummy. The regression of y on $(\bar{v}^{ob}p, p1_k, x, 1_s, z = 1)$ is studied.

This proposal leads to two practical questions. First, how many groups should be formed? That is, which K should be used? In simulations and empirical analysis, I present estimates based on a range of K from 1 to S : 1, 3, 15, 25 and 75. The unrestricted estimator has $K = S = 75$. Second, how should the schools be partitioned into groups? In theory, grouping schools with similar missing student means should reduce bias. If schools s and s' had equal

⁷ Consider replacing $\beta\bar{v}$ in the production function with the more general $h(\bar{v}) = h(\bar{v}^{ob}p + \bar{v}^m(1-p))$. What property of h delivers point identification of h with *MI1*? Linearity. Mathematically, it allows the expectation operator to pass inside the function and to separate the missing and observed components. Substantively, it says that variation in the observed portion is sufficient to identify a peer effect.

Consider studying the effect of other properties of the peer v distribution besides the mean, \bar{v} . A different issue arises. Although random assignment guarantees that the expectation of \bar{v} is constant within a school regardless of the number of missing-data students in a class, the same is not true for the expectation of other properties of the peer v distribution. For instance, the variance of mean missing peer v^m depends on the number of $z = 0$ students whose v^m are being averaged to form \bar{v}^m . Therefore, across classes with different numbers of missing-data students, the variance of \bar{v}^m will not be constant. This can be dealt with using observable information on the number of students missing but it requires a different approach.

missing data means ($E[v^m|s] = E[v^m|s']$), equation (5) shows that grouping them would create no bias and would reduce variance. However, $E[v^m|s]$ are never observed. Assumptions about the relationship between observables and $E[v^m|s]$ can be used. For instance, assume that the missing student mean and observed student mean have the same rankings across schools: $E[v^{ob}|s] < E[v^{ob}|s'] \Rightarrow E[v^m|s] < E[v^m|s']$. This suggests partitioning the schools into groups on the basis of their observed student mean v rankings — dividing them into K -ciles. If true, this should lead to less bias than some other partitions into K groups.

To understand the bias of these estimators, consider the limiting case where no group dummy $\times p$ terms are included ($K = 0$). Transforming to first differences permits a relatively simple expression for bias in this case. Define w^* as the difference between any variable $w = y, \bar{v}^{ob}p, p, e$ and its expectation conditional on $(x, 1_s, z = 1)$: $w^* \equiv w - E[w|x, s, z = 1]$. Then,

$$y^* = \beta \bar{v}^{ob}p^* + \beta E[v^m|s]p^* + e^*$$

The bias in estimating β by regression of y^* on $\bar{v}^{ob}p^*$ is the covariance of the regressor with the omitted terms $E[v^m|s]p^*$ over the variance of the regressor,

$$\beta \sum_s E[v^m|s] \frac{E[1_s(\bar{v}^{ob}p^*)(p^*)]}{V(\bar{v}^{ob}p^*)}. \quad (6)$$

In this case, bias increases in β , $E[v^m|s]$, and $E[1_s(\bar{v}^{ob}p^*)(p^*)]$. It decreases in $V(\bar{v}^{ob}p^*)$. This is analogous to a panel data model where there is a common main effect of an interaction of two variables and a unit-specific effect of one of the variables.

Monte Carlo simulations can compare the performance of p -weight estimators with various K to one another and to the two other estimators in the literature — the IDP and AP-corrected IDP — in terms of bias, variance, and mean squared error (MSE).⁸ We consider four different data-generating processes (DGP). In every DGP, observed data are taken as given, values for the missing v^m data are drawn for each $z = 0$ student from a school-specific distribution, and outcomes are generated. Then the generated v^m data are deleted and peer

⁸The IDP and AP-corrected IDP will be discussed analytically in Section 3.3; discussion of the simulation results regarding them is deferred until then.

effects are estimated. DGP vary along two dimensions: how the ordering of schools' missing student mean lagged achievement $E[v^m|s]$ was determined and whether teachers (correlated effects) are absent/present. For each DGP, 5,000 Monte Carlo replications were performed. In each replicate, all seven estimators were used to estimate β . The distribution of each estimator's 5,000 $\hat{\beta}$ s measure its squared error, variance, and MSE.

Monte Carlo results are presented in Table 1. Each estimator's MSE is expressed as a multiple of the MSE for the unrestricted p -weight estimator in the no-teacher condition under the same $E[v^m|s]$ -ordering process. The share of each estimator's total MSE due to squared error and to variance is also displayed.

What do the results tell us about the performance of the p -weight estimators with various K ? Consider the top left panel where the assumption that the missing student mean and observed student mean have the same rankings across schools ($E[v^{ob}|s] < E[v^{ob}|s'] \Rightarrow E[v^m|s] < E[v^m|s']$) is true. As expected, the unrestricted p -weight estimator appears unbiased. Squared error accounts for only 0.02 percent of the MSE whereas variance accounts for the other 99.98 percent. The MSE of the restricted p -weight estimators is less for all K . Bias does not appear to be a problem, but reducing the collinearity in the regressors and reducing the number of parameters estimated does reduce variance considerably. Including randomly assigned teachers raises the variance of all the estimators but does not affect bias. For instance, the MSE of the unrestricted p -weight estimator is 1.56 times larger when teachers are included as when they are excluded.

What happens if our assumption about the ranking of schools' $E[v^m|s]$ is violated? The right panel of Table 1 assigns $E[v^m|s]$ randomly while ensuring that they have similar distributional characteristics as obtained in the other DGP. Relative performance does not change much, although the variance of all estimators increases. The unrestricted p -weight estimator in the no-teacher condition has an MSE that is 17.4 times larger in the random ranking condition than in the $E[v^{ob}|s]$ ordering condition. The MSE decreases proportionally less as K decreases and increases more when teachers are considered. However, in every case, bias remains minimal. Shifting the $E[v^m|s]$ to extreme values does increase bias considerably for

$K < S$ but not for $K = S$, as expected from equation (6) (results available on request).

3.2.2 Analyzing students with missing x data

Thus far, we have ignored outcome and available covariate information for students with missing data on a peer-influencing covariate ($z = 0$). This section extends p -weight analysis to those with $z = 0$ using a few mild assumptions. Although some components of x are sometimes missing, other components are always observed. Divide the components of x between those that are sometimes missing (a) and those that are always observed (b) such that $x \equiv (a', b')'$.

We can use the observed $(y, b|z = 0)$ information to estimate β by studying the effect of observed peer mean on outcomes among those with $z = 0$. Although we cannot condition on a for these students, random assignment ensures that observed peer quality is mean independent of observed characteristics and therefore delivers an unbiased estimate of peer effects. With nonparametric f , no new structure is required. For linear $f(a, b)$, the expectation of a must be linear in b : $E[a|b, s, z = 0] \equiv \delta a$ (see note for details).⁹ We can combine the analysis done

⁹Let a new conditioning set be $\tilde{d} \equiv (\bar{v}^{ob}, p, b, s, z = 0)$. Assume two mean independence conditions closely linked to those defined earlier and one new one. These are formally justified in Section A of the Web Appendix.

$$\begin{aligned} MI1_b : & \quad E[\bar{v}^m | \bar{v}^{ob}, p, b, s, z = 0] = E[v^m | s] \\ MI2_b : & \quad E[e | \bar{v}^{ob}, p, b, s, z = 0] = E[e | s, z = 0] \\ MI3_b : & \quad E[f(a, b) | \bar{v}^{ob}, p, b, s, z = 0] = E[f(a, b) | b, s, z = 0] \end{aligned}$$

Their credibility depends on random assignment. $MI1_b$ follows directly from $MI1$. $MI2_b$ is justified by the same logic as $MI2$. Neither one's observed peer mean, fraction of peers observed, own characteristics, nor unobserved teacher characteristics affect either the expectation of one's missing peer mean \bar{v}^m or within-school unobserved influences.

$MI3_b$ is a different kind of condition. It requires that the expected effect of one's own (a, b) does not vary with the realized mean among one's observed peers \bar{v}^{ob} or the fraction p of observed peers. It deals with the expected effect of $f(a, b)$ on y where $f(a, b)$ is integrated over the distribution of (a, b) conditional on \tilde{d} . In the production function, (a, b) does not interact with \bar{v}^{ob} or p , so that the productivity of own (a, b) does not change with (\bar{v}^{ob}, p) . By random assignment, $P(a, b | \bar{v}^{ob}, p, b, s, z = 0) = P(a, b | b, s, z = 0)$. These imply $MI3_b$.

Taking the expectation of the production equation conditional on \tilde{d} gives

$$\begin{aligned} E[y | \tilde{d}] &= \beta \bar{v}^{ob} p + \beta E[\bar{v}^m (1 - p) | \tilde{d}] + E[f(a, b) | \tilde{d}] + \gamma_s + E[e | \tilde{d}] \\ &= \beta \bar{v}^{ob} p + \beta E[v^m | s] - \beta E[v^m | s] p + E[f(a, b) | b, s, z = 0] + \gamma_s + E[e | s, z = 0]. \end{aligned}$$

The first term identifies β . The covariates b no longer identify f . They pick up a mix of direct effects of b and indirect effects of a correlated with b . For instance, suppose f is additively separable in a and b : $f(a, b) \equiv f_a(a) + f_b(b)$. Then variation in b identifies $E[f_a(a) | b, s, z = 0] + f_b(b)$. The other terms work as in Section 3.2.1. For linear $f(a, b) \equiv \pi_a a + \pi_b b$, which is used in estimation, the conditional expectation of a must be linear in b , $E[a | b, s, z = 0] \equiv \delta a$, and the coefficient on b becomes $\pi_a \delta + \pi_b$.

on $z = 0$ and $z = 1$ students into one analysis. In practice, this is implemented in a single regression with appropriate interactions with z .

Thus far, we have not pursued the traditional route for dealing with missing data, which is to impose some equivalence across the distributions of (x, e) between students with $z = 1$ and $z = 0$ such as MCAR or MAR. Everything has been done in a way consistent with data missing not at random (MNAR). In the application, this approach seems appropriate. The following section proposes a new method for improving inference under stronger assumptions.

3.2.3 Improving inference with information about variance ratios

This section develops conditions that allow use of observed differences across classes in the missing-data students' outcomes (y) and never-missing covariates (b) to make inference about differences across classes in the average missing peer-influencing variables \bar{v}^m . This brings more information to bear in making inference about β . The variables used are the same as in the previous section, but now (y, b) conditional $z = 0$ are taken as informative about the class's \bar{v}^m , whereas the only structure imposed on \bar{v}^m until now was *MI1*. New structure is assumed, which tightens inference. The approach borrows from the measurement error literature [Boggs et al. (1987), Carroll (2006)] and exploits information about the ratio between the variance of ϵ for any $z = 1$ student and the variance of the average ϵ among all missing-data students in the same class. This information depends on the assumption that $z = 1$ and $z = 0$ students' ϵ come from distributions with the same variance conditional on observables. Such conditions provide a new way to identify peer effects in the presence of missing data that may be useful in many settings.

The study of peer effects provides a uniquely credible setting for using this variance ratio approach to measurement error because, here, one student's outcome error e is the source of other students' measurement error in covariates. In contrast, in most settings outside social interactions, measurement error derives from a source less closely related to the outcome error.

Recast the missing peer mean \bar{v}^m in terms of average observables and unobservables by adding up production functions across missing-data individuals and rearranging terms. Con-

sider any class- c . It has N_c students: n_c with $z = 1$ and $N_c - n_c$ students with $z = 0$. Define the average value among missing-data students in the class as $\bar{w}_{sc0} \equiv (N_c - n_c)^{-1} \sum_{i \in c} w_{sci}(1 - z_{sci})$ for any variable $w = y, a, b, p, \bar{v}^{ob}p, e$. This is a class average conditional on $z_{sci} = 0$, not a leave-me-out mean.

We use two new assumptions about functional form. First, assume production is linear ($f(a, b) \equiv \pi_a a + \pi_b b$). Second, assume peer effects operate through $a \equiv v = g(a, b)$; that is, we want to study the effect of the average value of the sometimes-missing, peer-influencing variable per se rather than a known function of it.

It follows that for any $z = 1$ student, her missing peer mean — the vexing variable that causes all the trouble — can be expressed as

$$\bar{v}_{sci}^m = \frac{\bar{y}_{sc0} - \beta \bar{a}^{ob} \bar{p}_{sc0} - \pi_b \bar{b}_{sc0} - \gamma_s - \bar{e}_{sc0}}{\beta(1 - \bar{p}_{sc0}) + \pi_a} \quad (7)$$

Details are available.¹⁰

¹⁰Consider any student- j with $z_{scj} = 0$. They all share the same values of variables relating to one's observed-data peers. The fraction of j 's peers with $z = 1$ is $p_{scj} = \frac{N_c - n_c - 1}{N_c - 1} = \bar{p}_{sc0}$, and the observed peer mean is $\bar{v}_{scj}^{ob} = \bar{v}_{sc0}^{ob}$. Their product is $\bar{v}_{scj}^{ob} p_{scj} = \bar{v}_{sc0}^{ob} \bar{p}_{sc0}$.

For any class- c , the $z = 1$ students' missing peer mean, which is the average $v = a$ of the $z = 0$ students in class- c , can be expressed as a function of observable characteristics of $z = 0$ students and other factors by manipulating the $z = 0$ students' production equations. The production model implies that, for any j ,

$$\begin{aligned} y_{scj} &= \beta \bar{v}_{scj}^{ob} p_{scj} + \beta \bar{v}_{scj}^m (1 - p_{scj}) + \pi_a a_{scj} + \pi_b b_{scj} + \gamma_s + e_{scj} \\ &= \beta \bar{v}_{sc0}^{ob} \bar{p}_{sc0} + \beta [(N_c - n_c - 1)^{-1} \sum_{k \neq j, z_{sck}=0} a_{sck}^m] (1 - \bar{p}_{sc0}) + \pi_a a_{scj} + \pi_b b_{scj} + \gamma_s + e_{scj}. \end{aligned} \quad (8)$$

Indexing the $z = 0$ individuals in class- c by $j = 1, 2, \dots, (N_c - n_c)$ and adding up their production equations (8) gives

$$\sum_j y_{scj} = (N_c - n_c) \beta \bar{v}_{sc0}^{ob} \bar{p}_{sc0} + \beta (1 - \bar{p}_{sc0}) \sum_j a_{scj}^m + \pi_a \sum_j a_{scj}^m + \pi_b \sum_j b_{scj} + (N_c - n_c) \gamma_s + \sum_j e_{scj}. \quad (9)$$

The a_{scj} in the $\pi_a \sum_j a_{scj}^m$ term are all missing since these are $z = 0$ students. They are the same values in the $\beta(1 - \bar{p}_{sc0}) \sum_j a_{scj}^m$ term. Any of the $z = 1$ students will have $\bar{v}_{sci}^m = (N_c - n_c)^{-1} \sum_{j \in c, z_{scj}=0} a_{scj}^m$. Solving equation (9) for this term yields

$$(N_c - n_c)^{-1} \sum_{j \in c, z_{scj}=0} a_{scj}^m = \frac{\bar{y}_{sc0} - \beta \bar{v}_{sc0}^{ob} \bar{p}_{sc0} - \pi_b \bar{b}_{sc0} - \gamma_s - \bar{e}_{sc0}}{\beta(1 - \bar{p}_{sc0}) + \pi_a}. \quad (10)$$

This can be substituted into the production function for $z = 1$ students in class- c and e expanded such that

$$\begin{aligned}
y_{sci} &= \beta \bar{v}_{sci}^{ob} p_{sci} + \beta \bar{v}_{sci}^m (1 - p_{sci}) + \pi_a a_{sci} + \pi_b b_{sci} + \gamma_s + e_{sci} \\
&= \beta \bar{a}_{sci}^{ob} p_{sci} + \beta \left[\frac{\bar{y}_{sc0} - \beta \bar{a}_{sc0}^{ob} \bar{p}_{sc0} - \pi_b \bar{b}_{sc0} - \gamma_s - \kappa_c - \bar{\epsilon}_{c0}}{\beta(1 - \bar{p}_{sc0}) + \pi_m} \right] (1 - p_{sci}) \\
&\quad + \pi_a a_{sci} + \pi_b b_{sci} + \gamma_s + \kappa_c + \epsilon_{sci}.
\end{aligned} \tag{11}$$

This single equation captures the relationship between all observable information given the model.

Straight forward estimation based on equation (11) runs into stumbling blocks. Consider the mean regression of y_{sci} on $\check{d} \equiv (\bar{a}_{sci}^{ob} p_{sci}, \bar{y}_{sc0}, \bar{a}_{sc0}^{ob} \bar{p}_{sc0}, \bar{b}_{sc0}, a_{sci}, b_{sci}, 1_s, z = 1)$. For each $z = 1$ student, there are two kinds of unobservables here: additive equation error $\kappa_c + \epsilon_{sci}$ and measurement error in the covariates expressed in the $-\kappa_c - \bar{\epsilon}_{c0}$ term.

By rearranging equation (7), the measurement error can be expressed as classical, additive error where the noisy measure (the LHS) equals a ratio of the latent true value (\bar{v}_{sci}^m) plus noise ($\kappa_c + \bar{\epsilon}_{sc0}$) that has conditional mean zero:

$$\bar{y}_{sc0} + \beta \bar{a}_{sc0}^{ob} \bar{p}_{sc0} + \pi_b \bar{b}_{sc0} + \gamma_s = [\beta(1 - \bar{p}_{sc0}) + \pi_a] \bar{v}_{sci}^m + \kappa_c + \bar{\epsilon}_{sc0}.$$

The measurement error and the equation error in equation (11) are correlated through κ_c . This causes problems. One could simply assume that no group-level unobserved influences are present: $\kappa_c \equiv 0$. This is not credible in the current application, but it is credible in some settings. If this assumption is maintained in error, estimates of β are upwardly biased. Correlation in y_{sci} and \bar{y}_{sc0} would be mistaken for the effect of peers rather than for the effect of teachers.¹¹

One approach to white-noise measurement error uses a restriction on the ratio of variances between the error in the outcome (ϵ_{sci}) and the measurement error in the regressor

¹¹Monte Carlo simulations are available on request.

($\bar{\epsilon}_{sc0}$) [Boggs et al. (1987), Carroll (2006)]. Assume that conditional within-class idiosyncratic influences for individuals in the same class all have the same variance and are independent across individuals:

$$V[\epsilon|\bar{a}^{ob}, p, a, b, c] = V[\epsilon|\bar{a}^{ob}, p, a, b, c, z]. \quad (12)$$

Given random assignment of individuals to classes from school populations, this would follow from $V[\epsilon|s] = V[\epsilon|z, s]$. Alternatively, it would also follow from the more general missing at random (MAR) assumption: $P[\epsilon|\bar{a}^{ob}, p, a, b, c] = P[\epsilon|\bar{a}^{ob}, p, a, b, c, z]$. Then, for any $z = 1$ individual in class- c ,

$$\frac{V[\epsilon_{sci}|\bar{d}]}{V[\bar{\epsilon}_{sc0}|\bar{d}]} = N_c - n_c. \quad (13)$$

The ratio of the outcome variance to the measurement error variance is exactly the number of missing-data individuals in the class, that is, the number of individuals over whom the average $\bar{\epsilon}_{sc0}$ is taken.¹² Averaging over more $z = 0$ students reduces the variance of $\bar{\epsilon}_{sc0}$ and raises the ratio. Under this condition, Boggs et al. (1987) develop an estimator based on nonlinear orthogonal regression. Adapted here, we propose to find the parameters that minimize,

$$\min_{\theta, \bar{\epsilon}_{sc0}} \sum_{z_{sci}=1} [(\hat{\epsilon}_{sci})^2 + (N_c - n_c)(\bar{\epsilon}_{sc0})^2],$$

where $\hat{\epsilon}_{sci}$ is defined as a function of data and parameters after solving equation (11) for ϵ_{sci} and $\bar{\epsilon}_{sc0}$ is a parameter shared by all $z = 1$ students in class- c . This can be implemented by Generalized Method of Moments. This approach, assuming $\kappa_c \equiv 0$, is implemented in the Section 5 but likely exaggerates peer effects.

3.3 Analysis of and comparison to existing approaches

Boucher et al. (2010) extend the analysis of Lee (2007) to deal with missing data when group-level fixed effects are used. They point out that these fixed effects absorb the influence of

¹²More generally, one could assume that $\lambda V[\epsilon|c, z = 1] = V[\epsilon|c, z = 0]$ with a known λ . The (equation error/measurement error) ratio would then be $\lambda(N_c - n_c)$.

missing-data peers as well as teachers or other contextual factors in analysis of the $z = 1$ sample. In the notation here, such a fixed effect would be $\gamma_c \equiv \kappa_c + \beta \bar{v}^m(1 - p)$. Rather than use across-group variation in peers, this approach identifies peer effects from within-group peer variation derived from leaving i out of the class mean v for different i . It uses a mean independence condition similar to MI defined at the class rather than school level. However, in the present application, this approach yields imprecise estimates as class fixed effects explain 99 percent of the variance in \bar{v}^{ob} .

The most common approach to missing data in the literature is to use an individual-deletion procedure (IDP). This approach produces biased and inconsistent estimates in the presence of missing data. In general the bias can be up or down or can even produce the wrong sign. When individuals are randomly assigned to groups, the bias attenuates estimates to zero. It acts similarly to white-noise measurement error, although it is not exactly that. The best approach in the literature is the correction to the IDP estimator proposed by Ammermueller and Pischke (2009). Its properties relative to p -weight estimators are explored analytically and through Monte Carlo simulation.

An IDP amounts to regressing y on $(\bar{v}^{ob}, x, 1_s)$ in the $z = 1$ subsample. For compactness, let the IDP conditioning variables be represented by a vector $d \equiv (\bar{v}^{ob}, x', s, z = 1)'$. An IDP leads one to study the following misspecified regression equation in the $z = 1$ subpopulation:

$$E[y|d] = \alpha \bar{v}^{ob} + f_{IDP}(x) + \alpha_s + E[e|d]. \quad (14)$$

The peer effect measured by IDP is how the expectation of y changes as \bar{v}^{ob} changes. The misspecification means that the coefficient on \bar{v}^{ob} is not constant. It varies with the particular values (\bar{v}^{ob}, x, s) at which the expression is evaluated. For this reason, it is denoted $\alpha(d)$:

$$\begin{aligned} \alpha(d) &= \frac{\partial E[y|d]}{\partial \bar{v}^{ob}} \\ &= \beta \left(E[p|d] + \bar{v}^{ob} \frac{\partial E[p|d]}{\partial \bar{v}^{ob}} + \frac{\partial E[\bar{v}^m|d]}{\partial \bar{v}^{ob}} - \frac{\partial E[\bar{v}^m p|d]}{\partial \bar{v}^{ob}} \right) + \frac{\partial E[e|d]}{\partial \bar{v}^{ob}} \\ &\equiv \beta \Lambda(d) + \frac{\partial E[e|d]}{\partial \bar{v}^{ob}}. \end{aligned} \quad (15)$$

The $\hat{\alpha}$ that is estimated using an IDP is a weighted average of this quantity across values of d . By equation (15), $\hat{\alpha} \rightarrow \beta$ only if $\frac{\partial E[e|\bar{v}^{ob}, x, s, z=1]}{\partial \bar{v}^{ob}} = 0$ and $\Lambda(d) = 1$ for all (\bar{v}^{ob}, x, s) such that $Pr(\bar{v}^{ob}, x, s, z = 1) > 0$. In general, $\Lambda(d)$ can take any value. The IDP estimator $\hat{\alpha}$ aggregates across estimates of $\alpha(d)$, none of which equal β .

Therefore, the IDP estimator is biased and inconsistent without a general direction, sign, or magnitude. The bias depends on the interaction of the missingness and group formation processes as expressed in $\Lambda(d)$. Asymptotically, the IDP estimate could be larger or smaller than the true parameter or of an opposite sign. Caution is in order when interpreting IDP estimates in the literature.

We now study how an IDP performs assuming that students are assigned randomly to classes. As before, random assignment implies that errors are mean independent of \bar{v}^{ob} ($\frac{\partial E[e|\bar{v}^{ob}, x, s, z=1]}{\partial \bar{v}^{ob}} = 0$) and helps separate causal from correlated effects. However, random assignment also ensures that $\Lambda(d) < 1$ and that the IDP estimator is attenuated to zero. Under random assignment of students to classes, any student's p will be independent of her $(\bar{v}^{ob}, \bar{v}^m, x)$ conditional on $(z = 1, s)$. This gives both $\frac{\partial E[p|d]}{\partial \bar{v}^{ob}} = 0$ and $\frac{\partial E[\bar{v}^m p|d]}{\partial \bar{v}^{ob}} = \frac{\partial E[\bar{v}^m|d]}{\partial \bar{v}^{ob}} E[p|d]$. Further, random assignment implies that the likelihood of any student's missing peer mean \bar{v}^m taking a particular value is independent of (\bar{v}^{ob}, x, z) conditional on s . This can be expressed as a mean independence condition that follows from *MI1*, one that excludes p from the conditioning set: $E[\bar{v}^m|\bar{v}^{ob}, x, s, z] = E[\bar{v}^m|s]$, which implies $\frac{\partial E[\bar{v}^m|d]}{\partial \bar{v}^{ob}} = 0$. Therefore,

$$\begin{aligned}
\Lambda_{RA}(d) &= E[p|d] + \bar{v}^{ob} \frac{\partial E[p|d]}{\partial \bar{v}^{ob}} + \frac{\partial E[\bar{v}^m|d]}{\partial \bar{v}^{ob}} - \frac{\partial E[\bar{v}^m p|d]}{\partial \bar{v}^{ob}} \\
&= E[p|d] + (1 - E[p|d]) \frac{\partial E[\bar{v}^m|d]}{\partial \bar{v}^{ob}} \\
&= E[p|\bar{v}^{ob}, x, s, z = 1] \\
&= E[p|s, z = 1] < 1.
\end{aligned} \tag{16}$$

The first equality is the definition of $\Lambda(d)$, the second comes from the independence of p and $(\bar{v}^{ob}, \bar{v}^m, x)$, the third from the definition of d and $\frac{\partial E[\bar{v}^m|d]}{\partial \bar{v}^{ob}} = 0$, and the last from the

independence of p from (\bar{v}^{ob}, x) conditional on s . The inequality comes from the presence of missing data in each class.

Random assignment of students within school makes $\Lambda_{RA}(d)$ invariant within school. Using equations (15) and (16), we can define $\alpha_{RA}(d) = \beta E[p|s, z = 1] \equiv \alpha_{RA}(s) < \beta$. Given finitely many students and classes per school while allowing the number of schools to go to infinity, the IDP estimator aggregates across schools with different $E[p|s, z = 1]$. No $E[p|s, z = 1]$ is identified. The IDP estimator of peer effects $\hat{\alpha}_{RA}$ aggregates across the sample. Given infinitely many classes in a school- s , β could be identified using only information from within school- s . As $n_s \rightarrow \infty$, $E[p|s, z = 1]$ would be observed and β would be identified from the IDP as $\alpha_{RA}(s)(E[p|s, z = 1])^{-1}$.

Monte Carlo simulations contrast the properties of the biased, precise IDP estimator and the unbiased, imprecise unrestricted p -weight estimator developed in Section 3.2.1. In these simulations, various fractions of the simulated a values are then censored.¹³ Peer effects are estimated using both IDP and p -weights. Results are displayed in Figure 2. The IDP estimates grow more attenuated as the fraction of missing data grows. The p -weight estimator's confidence intervals contain the true value for all degrees of missingness, but it is imprecise.

Ammermueller and Pischke (2009), through a different route of analysis, arrive at a similar finding and propose a correction to the IDP estimator. They suggest multiplying the IDP estimate by $C^{-1} \sum_c \frac{N_c - 1}{n_c - 1}$. Conceptually, this correction factor is very similar to $(E[p|s, z = 1])^{-1}$. Approximately, the AP correction scales up the IDP's point estimate and standard errors by the inverse of the fraction of data observed.

Returning to Table 1, we can compare the performance of the IDP and AP-corrected estimators with that of the p -weight estimators. This evidence illustrates the analytic points made above and shows that across a range of DGP, the restricted p -weight estimators have minimal bias and reduced variance. Together this translates into lower mean squared error than the alternatives. As expected, the IDP estimator is biased but precise. Its MSE is 51.85 percent that of the unrestricted p -weight estimator but larger than that of all the restricted

¹³In the Table 1 simulation, in contrast, all simulated a values were subsequently censored so that the fraction of missing data always matched that of the original data.

p -weight estimators. Unlike the p -weight estimators, squared error makes a large contribution to the IDP's MSE. The AP correction generally has a smaller MSE than the unrestricted p -weight estimator but larger than the MSE of any of the restricted versions. It also has a larger MSE than the IDP estimator. It is much less biased but also much less precise.

The problem with IDP estimates should not be confused with a classical errors-in-variables problem where the observed variable is the true variable plus mean zero noise. To see this, express the noisy observed variable \bar{v}^{ob} as a function of the true causal variable \bar{v} plus noise. Rearranging the decomposition of \bar{v} gives

$$\bar{v}^{ob} = \frac{1}{p}\bar{v} + \frac{1-p}{p}\bar{v}^m$$

Conditional on any true \bar{v} and any $p \in (0, 1)$, the expectation of \bar{v}^{ob} is

$$E[\bar{v}^{ob}|\bar{v}] = \frac{1}{p}\bar{v} + \frac{1-p}{p}E[\bar{v}^m|\bar{v}]$$

By definition, the noise is mean zero if $E[\bar{v}^{ob}|\bar{v}] = \bar{v}$. This requires $E[\bar{v}^m|\bar{v}] = \bar{v}$. Without some further restriction on the distributions of v^{ob} and v^m , this will not hold. Even if $E[v^m|s] = E[v^{ob}|s]$, finite sample variance ensures the equality does not hold generally since the realized $\bar{v} \neq E[v|s]$.

3.3.1 Partial identification with data not missing at random

We have explored how a condition on the peer group formation process, random assignment, gives point identification of β without requiring conditions on the data missingness process. The identifying power of the random assignment condition can be illustrated by relaxing it and seeing what we learn from the model and the observable data in its absence.

We can study the set of values of θ consistent with the observed data and the assumed model given any feasible distribution of the missing data. We translate ideas developed in Horowitz et al. (2003) to the present context. Considering all possible missingness processes and all possible group formation processes, the identification set $H[\theta]$ is the set of possible

production parameters θ given population data, $P(\bar{v}^{ob}, p, x, z, s)$.

Missing peer data imply that, for each individual, the researcher observes an interval known to contain the true peer mean \bar{v} . For any student, the bounds are determined by (\bar{v}^{ob}, p) and the logical bounds on $g(x)$. Denote the logical bounds as $v_- \equiv \min_{x \in X} g(x)$ and $v_+ \equiv \max_{x \in X} g(x)$. Any student's true peer mean is $\bar{v} = \bar{v}^{ob}p + \bar{v}^m(1 - p)$. Therefore, the true peer mean for a student with (\bar{v}^{ob}, p) must be between the lower and upper bounds defined as

$$\begin{aligned}\bar{v}_- &\equiv \bar{v}^{ob}p + v_- \cdot (1 - p) \\ &\leq \bar{v} \leq \\ \bar{v}_+ &\equiv \bar{v}^{ob}p + v_+ \cdot (1 - p).\end{aligned}$$

Focus on $v = x$, end-of-kindergarten achievement percentile ($v_- = 1$ and $v_+ = 99$).

This implies an identification region for θ . This region can be described abstractly, and computing it is feasible. However, it is quite burdensome computationally. Inference on mean regressions with fully observed outcomes and interval-observed covariates is much more difficult than interval-observed outcomes and fully observed covariates.¹⁴

One can learn about $H[\theta]$ by identifying elements that belong to it. These known members constitute an inner bound on the set, $C \subseteq H[\theta]$. Elements $H[\theta]$ can be identified by plugging in various feasible values for the classrooms' mean missing-data students and computing the parameters of interest θ . The mean v among those with missing v in class- c is $\bar{v}_c^m \in [v_-, v_+]$, $\forall c = 1, 2 \dots C$. Let \vec{v}^m represent a particular point in the $[v_-, v_+]^C \equiv M$ space of feasible values. Given observed data and the production model, each \vec{v}^m implies a value for the unknown θ (other than $P_{e|\bar{v}^{ob}, \bar{v}^m, w, s, z, p}$) and, in particular, for β . Given any \vec{v}^m , a fully observed $\bar{v} = \bar{v}^{ob}p + \bar{v}_c^m(1 - p)$ follows for each $z = 1$ student in class- c . With “complete” covariate data for all $z = 1$ students, θ is identified. Therefore, the implied

¹⁴For intuition, consider a simple univariate regression estimator $\hat{\beta} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$. Compare the case of interval observation of a single x_i to interval observation of a single y_i . It is not trivial to understand how $\hat{\beta}$ changes as y_i takes different values in an interval because y_i influences \bar{y} and hence ripples back through every observation's contribution to the sum in the numerator. However, understanding how $\hat{\beta}$ changes as x_i takes different values in an interval has that same problem compounded by the fact that it occurs in both the numerator and the denominator.

production parameters can be considered a function of the missing value parameters, $\theta(\vec{v}^m)$. The identification set $H[\theta]$ is the set of $\theta(\vec{v}^m)$ for all $\vec{v}^m \in M$. An inner bound on $H[\theta]$ is identified by repeatedly computing $\theta(\vec{v}^m)$ for any set of \vec{v}^m . The inner bound C is the resulting set of $\theta(\vec{v}^m)$, $C(\tilde{M}) \equiv \{\theta(\vec{v}^m) | \vec{v}^m \in \tilde{M} \subseteq M\}$.

For any particular production parameter θ_k , the minimum and maximum of vectors in the inner bound merit special attention. They estimate the range of parameter values consistent with the data. For any inner bound C and any of its dimensions $k = 1, 2, \dots, \dim(\theta)$, define C_k^- as the minimum value in dimension k among the vectors belonging to C and C_k^+ as the maximum. By the definition of inner bound, we know that $\exists a, b \in H[\theta]$ such that $a_k \leq C_k^-$ and $b_k \geq C_k^+$. Given the observed data and the model, there exist distributions of missing data consistent with β less than or equal to the lower inner bound and greater than or equal to the upper inner bound.

To estimate an inner bound, we combine the observed sample with a set of feasible \vec{v}^m . The set of feasible values can be chosen in many ways. Here, an optimization routine is used to search over the space of feasible \vec{v}^m . To estimate the inner bound on the k^{th} -component of θ , the routine maximizes and minimizes $\theta_k(\vec{v}^m)$ with respect to \vec{v}^m .

4 Data

4.1 Sample

Tennessee’s Student/Teacher Achievement Ratio Project (Project STAR) tested the effects of small versus regular versus regular-with-aide classes on student achievement in grades K through 3. Word et al. (1990) and Krueger (1999) provide detailed discussions of the design, implementation, and results for class size effects. Generally, they found positive average achievement effects of small classes. Small violations of STAR’s random-assignment-to-class-type design have long been apparent. Some show up in the sample studied here. For instance, 2.2 percent of students in nonsmall first grade classes attended small kindergarten classes (see Table 11). The design says there should be none. The literature suggests that deviations from the design were minor [Krueger (1999), Hanushek (1999), Nye et al. (2004)]. Hanushek raises concerns about attrition undermining the internal validity of multiyear, class size effect estimates. The critique does not apply to the present study, which studies only a single year. It would apply to attempts to study the effect of first grade peers on later outcomes that suffer from attrition.

Generally, no pre-assignment measure of student achievement is available in STAR. When students entered a STAR school for the first time, they were assigned to a class and took achievement tests only at the end of each school year. Each student was supposed to stay in the same class group for the study’s duration.

Without pre-assignment measures of achievement, previous peer-effects researchers using STAR data have not focused directly on peer achievement and conditional prediction. They focused either on simultaneously determined achievement outcomes [Boozer and Cacciola (2001), Graham (2008)] or on peer effects defined in terms of pre-assignment covariates other than achievement [Dee (2004), student-teacher racial match, Whitmore (2005), peer gender; Cascio and Schanzenbach (2008), peer age].

However, Project STAR re-randomized returning first grade students who had been in either regular or regular-with-aide kindergarten classes to new classes of one of these two types. STAR also randomly assigned teachers and newly entering students to first grade

classes. Therefore, for returning first grade students, the data have gold standard features for separating causal peer effects from correlated effects: pre-assignment measures of achievement, random assignment to peers and teachers, and end-of-year outcomes. This re-randomization has been noted in passing [Word et al. (1990), Krueger (1999), Nye et al. (2000), Finn et al. (2007)] but never before exploited. Section 4.3 explores in depth whether they were randomly assigned only to a class condition (small, regular, regular-with-aide) or to specific classes within condition.

The complicating issue is that we do not observe a lagged achievement score for about 40 percent of the first grade students. Tennessee did not mandate kindergarten and many students entered school for the first time in first grade. Others may have been repeating first grade or may have come from a non-Project STAR school’s kindergarten. These students with missing peer-influencing covariate data appear in every school. The fractions vary significantly across schools.¹⁵ Like returning students, these missing-data students were also randomly-assigned within school. There is variation in the fraction of students with missing data between classes within school, consistent with small-sample variation under random assignment. Every first grade class contains more than one student with missing kindergarten achievement data. The STAR data include 215 first-grade, nonsmall classes in 76 schools. This paper analyzes students in 210 classes in 75 schools.¹⁶

4.2 Measures

The following section describes the measures of $(y, x, v, z, p, \bar{v}^{ob}, s)$ used. Table 2 describes each variable briefly and is intended for easy reference. Table 3 presents summary statistics.

Outcome y . Project STAR administered a battery of achievement tests to students at the end of each academic year. In first grade, these included the Stanford Achievement Test (SAT) in mathematics, total reading, and listening to words and stories [Word et al. (1990)]. In each

¹⁵ Across the 76 STAR schools, the fraction of first-grade students in nonsmall classes with lagged achievement observed ranges from 26 to 86 percent with a median of 59 percent.

¹⁶ In two schools with more than 3 classes, a single class was dropped because fewer than two students in the class had kindergarten achievement observed. One school was dropped because it contained only a single nonsmall class, so no within-school, between-class variation was possible. Two other classes in multiple classroom schools were dropped because all students’ low-income status was unobserved.

domain, students' performance can be scaled many ways [Gardner et al. (1982), Gardner (1983)]. Within domain and across years, the scale score is supposed to give a linear scale of achievement. However, scales are not equivalent across domains, and every scale itself is arbitrary. A unit gain does not mean anything in particular.

In order to obtain a meaningful scale and comparability across domains, scale scores in each domain are converted into nationally normed achievement percentiles. Specifically, each student's achievement level in the spring of first grade is measured as the average of her achievement percentiles on the math, reading and listening skills tests.¹⁷

Own covariates x . One of the unique and valuable features of the current analysis is the availability of pre-assignment, lagged achievement measures for most individuals. Each student's lagged achievement measure is constructed in the same way as y but with kindergarten rather than first grade scores. Call this end-of-kindergarten achievement percentile, x_1 . Correlations between math, reading, and listening skills percentiles and their average, which is the variable analyzed, in kindergarten and first grade regular-sized classes is presented in Table 12.

For each student's own covariates x , we also observe demographic characteristics from which we construct the following variables. We analyze indicators of female gender, African-American race, and low family income measured by each student's qualification for free or reduced-price lunch in first grade. Over 99 percent of the students are African-American or Caucasian. Students' birth dates are transformed into two variables: a continuous years of age measure and an indicator of whether the student was old enough to have attended first grade the prior year. We do not know why each older student is there. Presumably, they either entered school late at their parents' discretion or repeated a grade.¹⁸

¹⁷If not all three scores were observed, the scores were assumed to be missing completely at random and the average among the student's observable scores was used. Exploratory analysis suggested that the results did not differ when various combinations of dimensions were analyzed. Word et al. (1990) used nationally normed percentiles in the original analyses of Project STAR class size effects. However, the publicly-released data from Project STAR contains only the raw and scale scores. So, subsequent analyses measured achievement only standardized *within the STAR sample* [Krueger (1999), Krueger and Whitmore (2001), Dee (2004), Whitmore (2005), Cascio and Schanzenbach (2008), Graham (2008)]. Hopefully, the nationally normed percentile units used here are easier to interpret.

¹⁸A few students are missing demographic information. The pattern of missing data on y , x , and all other demographics is in Table 13. This model is built to handle the first two patterns of missing data: fully observed

Peer-influencing variables $v = g(x)$. One’s outcome y depends on the mean of first grade classmates’ kindergarten achievement percentile x_1 and so $v = g(x) \equiv x_1$. For any student- i in class- c , the causal peer variable is the mean of peers’ kindergarten achievement percentile, $\bar{v} = (N_c - 1)^{-1} \sum_{j \in c, j \neq i} x_{1j}$. Some specifications include peer demographic variables. These are the fraction of peers African-American, fraction female, and fraction low income. Other specifications study the impact of fraction of classmates in each achievement tercile, rather than the effect of classmates’ mean achievement percentile.

Observability indicator z . Observable kindergarten achievement level x_1 is indicated by z . For $z = 0$ students, x_1 is missing. In the notation introduced in Section 3.2.2, $a = x_1$. All other components of x equal b , including the fully observed peer demographic variables.

School attended and classroom peers with observed data, (s, p, \bar{v}^{ob}) . Each student’s STAR school and classroom teacher each year are observed. We analyze a vector of school indicators, 1_s . With the classroom-teacher identifiers, we define each student’s peers as his or her first grade classmates. Linking specific teachers and students to classrooms is valuable for understanding the most relevant influences on student outcomes.

Some studies of peer effects in schools using other data cannot connect students to specific classrooms or teachers. For instance, Zabel (2008) has classroom identifiers for students but not teachers. Hoxby (2000), Hanushek et al. (2003), and Hoxby and Weingarth (2007) can observe each student’s and each teacher’s school and grade but not their specific class. Betts and Zau (2004) compare the relative importance of classroom and school-grade peer cohorts for academic achievement outcomes and conclude that the classroom peers are more important. The ability to link each student to classroom peers and to a particular teacher in the STAR data has value.

Based on each student’s z and her classroom identifier, we can compute the fraction of classmates with observed v (p) as well the average $v = x_1$ among the student’s peers with observable v , $\bar{v}^{ob} = \bar{x}_1^{ob}$.

(y, x) or only x missing. The x values of the third type (only y missing) are used in computing their peers’ \bar{v}^{ob} , but these students’ own outcomes are not analyzed. Only 28 students attended first grade but have both (y, x) missing. For the purposes of computing p , they count as missing v . Students in the final group all have some missing individual information such as race, gender, age, or family income. The outcomes of individuals in the last two groups are not analyzed.

Class-level variables, part of x . A set of class-level variables is constructed and included in x in some specifications. These include an indicator of whether the class is regular type. Regular-with-aide is the omitted category. This is useful since regular-with-aide classes achieve a 1.8 percentile advantage over regular classes at the end of first grade. This occurs even though there was no significant achievement difference between the types at the end of kindergarten. Further, there was no significant difference in x between the first grade nonsmall class types, consistent with effective re-randomization at the start of first grade.

Small classes are not in the sample primarily because these students were not re-randomized at the beginning of first grade. Also, by focusing only on regular and regular-with-aide classes, we condition on a narrow range of regular classes sizes. However, Krueger (1999) analyzes the effect of within-class-type variation in class size on achievement percentile. He finds that the effects of class size are linear even within type. Therefore, we condition on class size to deal with variation in class sizes within the nonsmall classes.

Three variables capturing teacher quality are also included in specifications with class-level variables. First, the data include each teacher’s number of years of experience. Hanushek and Rivkin (2006) argue that returns to experience flatten out after the first year or two. Boyd et al. (2006) find that returns increase only for three to five years. In the present analysis, results do not change when a dummy for new teachers is included. Two other indicators capture whether each teacher has at least a master’s degree and is on a step of the career ladder (rather than being an apprentice, on probation, or not-on-ladder). All STAR teachers have at least a bachelor’s degree.

4.3 Evidence on student and teacher assignment processes

To achieve point identification, this paper assumes that students and teachers were randomly assigned, not just to a class type (small, regular, regular-with-aide) but to a specific class within type. Random assignment of teachers to classes guards against the possibility that teacher quality covaries with peer quality. Random assignment of students guards against the possibility that own or peer unobserved characteristics covary with a student’s own or peer

observable characteristics.

Although the design of Project STAR and its technical documentation make clear that assignment to class type was intended to be random, they do not speak to the process by which students or teachers were assigned to specific classes within type. Project STAR was designed to study class size effects, requiring random assignment to class type but not necessarily to specific classes. The Project STAR technical report [Word et al. (1990)] and the public-use data manual [Finn et al. (2007)] make clear that assignment to type was random. They are silent on the assignment process to classes within type. Articles analyzing class size effects with the STAR data do not go into detail about assignment to classes. However, Zaharias et al. (1995), Finn and Achilles (1999), and Finn et al. (2001) all suggest random assignment of students and teachers to specific classes within type.

Researchers who have previously used the STAR data to study teacher or peer effects refer to this ambiguity about the process of assignment to specific classes. For instance, Cascio and Schanzenbach (2008) state, “Project STAR program documents indicate no clear direction about how students were allocated when there were multiple classrooms of a given size within a school.” Some used analytic strategies that avoided assumptions on the within-type assignment process [Boozer and Cacciola (2001), Graham (2008)]. Others probed the data for evidence of nonrandom assignment of students and teachers to classes within type by looking for systematic differences in covariates across classes, replicating analyses in a subsample with more certain random assignment, or replicating analysis with instruments for class assignment [Dee (2004), Nye et al. (2004), Whitmore (2005), Cascio and Schanzenbach (2008)]. They all conclude that deviations were likely small and proceed with analysis assuming random assignment to classes.¹⁹

I also personally asked original Project STAR investigators to explain the process of student and teacher assignment. According to Charles Achilles, the STAR investigators took each school’s list of teachers and new students, randomly assigned them to specific classes within

¹⁹In *Econometrica*, Graham (2008) goes beyond expressing ambiguity about assignment within type and states, “Random assignment, for both students and teachers, was to class type *not* to specific classrooms.” He does not provide evidence for this claim. In personal communication on this issue, Graham stated, “My sense is that deviations from random assignment [to classes] were small.”

type, gave the rosters back to the principals, and monitored compliance through frequent site visits during each year. Jayne Boyd-Zaharias independently gave a consistent account. They have no incentive to falsely claim random assignment to classes within type, as it would not affect the validity of STAR’s class size results.

Further, I contacted many of the above-mentioned researchers — Boozer, Cacciola, Graham, Hedges, Konstantopoulos and Whitmore Schanzenbach — to ask what led them to doubt that random assignment happened at the class level. Did they know of any evidence of nonrandom assignment within type (or between types)? Uniformly, they pointed to the ambiguity in the documentation and the small deviations discussed above. None offered other specific evidence of nonrandom assignment within type.

Some peer-effects researchers who analyze administrative data where there is nonrandom assignment of students and teachers to classes look for balance in covariates across classes [Ammermueller and Pischke (2008), Graham (2005), Vigdor and Nechyba (2007)]. Previous analyses of STAR data have generally found that there is balance on observable covariates across classes. Nye et al. (2004) give evidence for each grade across all class types. Cascio and Schanzenbach (2008) and Graham (2008) study kindergarten only. We can study the data directly for evidence about the student and teacher assignment process in the main subsample of interest here, first grade students in nonsmall classes. As mentioned earlier, no other researcher has exploited the re-randomization at first grade so they all studied other samples. Across a variety of tests, we find strong but not absolute evidence of random assignment of students to classes. Graham (2008) find some statistical evidence that there is more gender mixing than would be expected from random assignment alone and we obtain similar results. Interested readers can find the results of a series of statistical tests of the null hypotheses of random assignment of students and teachers to classes in Section B of the Web Appendix.

5 Results

Analysis of the STAR data suggests that being assigned first grade classmates with higher lagged achievement significantly raises one’s own first grade achievement. This result emerges across a range of analyses premised on random assignment of students and teachers to specific classes and homogeneous peer effects. Although the unrestricted p -weight estimator proposed in Section 3.2.1 does not generate precise results, the restricted versions and the AP-corrected IDP yield more precise, significant estimates. Allowing for heterogeneous effects of peer mean achievement, there is some evidence that lower-achieving students benefit more than higher-achieving students from an increase in mean peer lagged achievement. However, this evidence is more tentative.

The identifying power of the random assignment condition in dealing with the missing-data challenge is illuminated by relaxing it. Allowing the missing data to take any distribution on its support (which corresponds to allowing any peer group formation process as well as any missingness process), we find that the observed data and the model are consistent with a very wide range of peer-effect parameter values. There exist distributions of the missing data that imply peer-effect estimates $\hat{\beta}$ below -0.45 and above 20.11. This illuminates an area of uncertainty created by the missing peer data and suggests the identifying power of the *MTI* condition. It also suggests how big a problem missing data can create in the context of peer effects.

5.1 Point identification of β assuming random assignment

Throughout this analysis, four basic specifications are used, each of which adds more covariates. Theoretically, they correspond to different definitions of x .

What do the specifications have in common? First, all include school fixed effects to deal with nonrandom selection into schools. Therefore, all identifying variation comes within schools. Second, all use end of first grade achievement y as the outcome. Later outcomes are less likely to be affected by first grade peers and also have higher fractions of data missing. Third, all standard errors allow for correlation of errors within classes.

How are the four specifications different? In addition to school indicators, specification A adds only a measure of mean peer achievement and own-achievement x_1 as predictors.²⁰ Specification B adds other individual, own-demographic variables: years of age and indicators for African-American, female, low-income family, and old-for-grade. Specification C adds first grade class-level variables: class size, teacher’s years of experience and indicators for teacher has a master’s degree, teacher on career ladder, and regular class (regular-with-aide omitted). Specification D adds three variables measuring the fraction of each student’s classroom peers who are African-American, the fraction female, and the fraction low-income. Here, demographic variables, not only lagged achievement, influence peers. However, the emphasis in the analysis remains on the effect of peer achievement conditional on these variables.

What effects of mean peer lagged achievement have other researchers found?²¹ Evidence has emerged across many settings and modes of analysis that, on average, higher-achieving peers have a moderate, positive effect on own achievement [Hanushek et al. (2003), Betts and Zau (2004), Cooley (2006), Hoxby and Weingarth (2007), Burke and Sass (2008), Graham (2008)]. Others have found negligible average effects [Lefgren (2004), Vigdor and Nechyba (2007), Zabel (2008)]. None studied the effects of peer achievement as early as first grade.

Table 4 presents results using the unrestricted p -weight estimator from Section 3.2.1. Here, mean lagged achievement among peers with lagged achievement observed \bar{v}^{ob} is weighted by p , the fraction of peers this includes. The coefficient on this product $\bar{v}^{ob}p$ estimates the peer-effect parameter β . Further, in addition to indicators for each school, this specification includes a set of interactions between each student’s fraction of peers observed and school indicators ($p \cdot 1_s$). These correspond to the $-\beta E[v^m|s]p$ term in equation (5).

The estimates of peer effects are imprecise and not statistically different from each other or from zero. In specification D, $\hat{\beta} = 0.12$ with a 95 percent confidence interval (CI) of

²⁰If own x_1 were omitted, then a negative correlation between the observed leave-me-out peer mean and the error, containing x_1 , would be induced. The estimated $\hat{\beta}$ would capture both true peer effects and the correlated negative of own- x_1 effects.

²¹The following discussion focuses on studies that model peer influence primarily via peer achievement (either lagged or simultaneous) rather than via demographic characteristics alone. This approach is somewhat controversial. Hoxby and Weingarth (2007) argue that, conditional on peer achievement, peer demographics do not have large effects although their interaction may be significant. In contrast, Vigdor and Nechyba (2007) conclude that peer demographics matter but peer achievement conditional on peer demographics does not.

$[-0.30, 0.54]$. This is consistent with peer effects ranging from moderately negative to strongly positive. Coefficients on own-demographics and class-level variables are plausible and precise suggesting the problem is not the specifications themselves.

Table 5 presents estimates of β from restricted p -weight estimators and the AP-corrected IDP estimator using the same sample and A-D definitions of x as in Table 4. Each cell presents a $\hat{\beta}$ from a separate regression. The columns give specifications A-D. The rows represent different estimators. These alternative estimators give more stable, consistently positive estimates of peer effects. In specification D, estimates range from 0.23 with a CI of $[-0.04, 0.51]$ to 0.37 with confidence interval $[0.01, 0.73]$. Here it is apparent that the unrestricted p -weight estimator, with estimates reproduced in the top row for reference, is an outlier to some extent. Across specifications A-D, unrestricted estimates are less stable than the restricted estimates and are generally lower and less precise.

How big would a $\beta = 0.25$ peer effect be? It implies that increasing peer mean kindergarten achievement by one nationally normed percentile raises own achievement by 0.25 percentile point on average. In the literature, it is common to standardize an estimated peer-effect parameter $\hat{\beta}$ into an effect size by multiplying $\hat{\beta}$ times $\frac{\sigma_v}{\sigma_y}$. This is especially important when the underlying variables have no meaningful scale. However, in this case, scales are meaningful. Both y and x_1 are in nationally normed percentile units. If we instead scale by $\frac{\sigma_v}{\sigma_y}$ and use the standard deviation of the national population rather than the STAR sample, then $\frac{\sigma_v}{\sigma_y} = \frac{\sigma_{x_1}}{\sigma_y} = 1$ and $\hat{\beta}$ can be interpreted directly as an effect size. For the national population with $P(y) = P(x_1) = 1/99$ for all $y, x_1 = 1, 2 \dots 99$, $\sigma_y = \sigma_{x_1} = 28.7$.²²

Thus far, we have only analyzed the $z = 1$ subsample. However, for the $z = 0$ students, we have (y, \bar{v}^{ob}, p) and all components of x other than kindergarten achievement. We can bring this information into the analysis as well. We begin with the analysis described in Section 3.2.2.

Table 6 presents the unrestricted p -weight results using outcomes from all students. These specifications interact own-demographic variables, class-level variables, and school fixed effects

²²Interested readers can find standard deviation σ of these sample variables in Table 17 of the Web Appendix. Within-school σ are calculated on school-centered variables.

with z . Coefficients on peer achievement \bar{v}^{ob} , peer demographics, and $p \cdot 1_s$ are the same regardless of z . Across specifications, the estimates of $\hat{\beta}$ are stable, ranging from 0.26 to 0.31. However, they are still imprecise, with CIs ranging from -0.15 below to 0.70 above.

Table 7 presents estimates of β from the restricted p -weight and AP-corrected estimators using the same sample and specifications as in Table 6. These estimated peer effects using these alternatives are larger, stable, and consistently significant. Across K , the restricted p -weight estimates in specification D are all between 0.34 and 0.38 with confidence intervals ranging from 0.04 below to 0.68 above. The AP-corrected IDP estimator yields even larger estimates.

Again, we see the pattern that as one moves from the unrestricted to the restricted p -weight to the AP-corrected estimators, the point estimates and significance increase. The formal analysis and Monte Carlo results suggest that this is not a general property of the estimators but a function of the specific data.²³

Why are the estimates of β larger in the full sample than the $z = 1$ sample? It could be that peer effects are stronger among kids entering school for the first time; recall that the $z = 1$ students attended kindergarten the prior year while the $z = 0$ students generally did not. Peers may matter more when you are learning how to be a student not just learning as a student. We cannot test this directly.

A different possibility, which we can test, is that (A) peer effects are heterogeneous depending on one's own achievement level and that (B) the distribution of $z = 0$ students is more heavily concentrated at lower achievement levels than the $z = 1$ distribution. There is evidence supporting (B). Compared to the $z = 1$ students, the $z = 0$ students are more likely to be African-American and low-income. They also tend to score lower at the end of first grade. Table 14 presents mean differences between populations on a variety of variables. Also, note from Table 3 that, among $z = 1$ students, only 17 percent scored in the bottom tercile of the nationally-normed achievement distribution, 31% scored in the middle tercile and the

²³Estimates based on assuming a known variance-ratio and no teacher effects as described in section 3.2.3 are presented in Table 8. These are not presented as credible point estimates but to make two points. First, the estimates are more precise than those considered earlier and even more precise than corresponding IDP estimates. The additional structure buys a lot of precision. Second, the point estimates of peer effects are larger. Since these mingle true peer effects with teacher effects, peer effects are no bigger than these.

remaining 52% scored in the top tercile. By definition, we do not know the achievement level of the $z = 0$ students but more likely would score in lower terciles. Is there evidence of (A), that peer effects vary by own achievement level?

Heterogeneous effects. Previous research suggests that the effect of the peer mean is not homogeneous. Hoxby and Weingarth (2007) use a flexible specification to investigate heterogeneous effects and to develop evidence on specific theories of how peer effects operate. For students in each decile, they study the effect of the fraction of (school-grade-year) peers in each decile. They reject a model of homogeneous effects of peer mean in favor of richer models that allow interactions between own and peer achievement.

Others researchers find that effects of peer mean achievement decrease with own achievement level. In Burke and Sass (2008), peer mean achievement has larger effects on students in the bottom quintile than the middle three quintiles. Effects are smaller still for top-quintile students. Hanushek et al. (2003) finds a reduced peer-mean-achievement effect in the top quartile. Ding and Lehrer (2007) and ? also find evidence consistent with diminishing peer-mean-achievement effects as own-achievement increases. If true, achievement mixing might raise average achievement and decrease achievement inequality compared to tracking.

Returning to the point-identification conditions, we look for heterogeneous effects of the peer mean. Here we allow the effect of peer mean to vary by whether one's own kindergarten achievement level is low, medium or high. The analysis of Table 4 and 5 is now redone interacting every predictor with a set of indicators for whether own x_1 is in the first, second or third tercile nationally, i.e. $1(x_1 \leq 33.3)$, $1(33.3 < x_1 < 66.7)$ or $1(66.7 \leq x_1)$.

Results from the unrestricted p -weight estimator appear in Table 9 on page 64. Estimates are imprecise. Specifications A-D are displayed in columns. Each row describes $\hat{\beta}$ for a different tercile. These are constructed as a main effect of $\bar{v}^{ob}p$ and the sum of the main effects and interactions between $\bar{v}^{ob}p$ and dummies for second and third terciles. Each coefficient can be distinguished neither from zero nor from coefficients from the other terciles at conventional levels. However, the point estimates are declining in own kindergarten achievement tercile. Whereas students who scored in the bottom tercile of kindergarten achievement experience an

estimated peer effect of 0.34, those effect for those who scored in the top tercile is 0.24.

Stonger but still suggestive evidence of heterogeneous effects from estimates of specification D under the alternative estimators are presented in Table 10, with the unrestricted reproduced in the top row for reference. Each row represents results from a separate regression with the listed estimator. The columns present the tercile-specific β_T estimates. The point estimates generally diminish as own achievement increases. For bottom-tercile students, the point estimates range from 0.34 to 0.84 and all are significantly different than zero for $K < S$. For students who scored in the top tercile of achievement, the point estimates range from 0.05 to 0.35. Only the largest is significantly different than zero. This evidence is not decisive. Only in the AP-corrected case are the top and bottom estimates significantly different than one another.

If these point estimates were taken seriously, they would imply that a policy of mixing students of different abilities in the same classes would tend to produce higher average achievement than would a policy of tracking. Gains among students in the bottom tercile due to mixing with middle- and high-achieving students would be greater than declines among the middle- and high-achievers from mixing with students in the bottom tercile. Also, the average outcomes of low achievers would be closer to those of high achievers, meaning reduced inequality in average outcomes. However, this analysis ignores the possibility that peer heterogeneity itself has effects on outcomes.

To deal with missing data, the current study requires peer effects to affect outcomes via a variable that is additively-separable in each individual peer's achievement. Neither standard deviation, interquartile range, nor other measures of spread in the peer distribution are so separable. We are not able to investigate the effect of spread *per se*.

The model does allow investigation of the effect of fraction of peers in a given set, such as tercile or decile, as done in Hoxby and Weingarth (2007) and Burke and Sass (2008). Burke and Sass (2008) analyze the effect of the proportion of classmates in the bottom, middle and top quintiles on students in each group. They conclude that maximizing average achievement would require mixing middle and high achievers and leaving low achievers separate. The

estimates we obtain from similar analysis are very imprecise.

Thus far, we have relied on assuming students were randomly-assigned to specific classes within class-size condition (regular versus regular-with-aide). There is more certainty that students were assigned randomly to condition. The internal validity of the Project STAR study of class-size effects hinges on random assignment to condition, but not to specific classes within condition. This suggests a check on our results thus far.

5.2 Partial identification

If we are willing to make due without point-identification, we can relax the assumption that students are randomly-assigned to classes and partially-identify the production parameters. We will still require that unobserved influences are uncorrelated with observed own, class and peer covariates. This may be a less credible condition now that we suppose students may not be randomly-assigned to classes. For instance, if students who were in the same classes in kindergarten share unobserved influences that emerge in first grade and still share a class for first grade, this would bias our estimates of peer effects [Hanushek et al. (2003)].

We estimate C_k^- and C_k^+ from inner bounds on $H[\beta]$ obtained by the procedures described in section 3.3.1. We search across values of the vector of missing peer means for each class in order to maximize and minimize $\hat{\beta}$. This generates peer-mean parameters that range from $\hat{C}^{M-} = -0.45$ to $\hat{C}^{M+} = 20.11$. We analyze the same 2,684 student $z = 1$ sample and include all additional covariates from specification D. Given the missing \bar{v}_c^m , we use $\bar{v} = \bar{v}^{ob}p + \bar{v}_c^m(1-p)$ as the regressor. Terms for $\bar{v}^{ob}p$ and $p \cdot 1_s$ are not needed.²⁴

This analysis shows that without making assumptions about the distribution of missing data, the observed data are consistent with a wide range of peer effects. Within the range are values that would imply radically different optimal policies if they were known to be true. And this is not a sharp inner-bound. The true range is possibly wider.

²⁴We have not dealt with the statistical uncertainty of this estimate of the inner-bound. Since we are estimating points on the boundary of the set, estimates are biased in finite samples. ? develop a bias correction based on a bootstrap heuristic. It lacks a formal theoretical foundation but may be useful in the absence of alternatives. Research on this and closely-related topics is proceeding but has not produced results yet.

6 Conclusion

The problems of inference on peer effects given missing data on some individuals' peer-influencing covariates deserves more attention. Empirical, peer-effects researchers commonly face this challenge but have few theoretical and methodological tools available to deal with it. The properties of one common approach, ignoring individuals with missing data and analyzing the observed-data students as if they represented the population, have not been well understood. As this paper points out, these IDP estimates of peer effects are biased and inconsistent. This paper gives an expression for how bias in IDP estimates of mean peer effects depends on the interplay between the missingness and the peer group formation processes.

Missing peer data are, in a sense, just another kind of unobserved influence. The challenge is to deal with the influence in a principled way that is consistent with one's theoretical model and is credible given what one knows about the data-generating process. The theorized structure of peer groups and of peer effects dictates how missing peer data matter for outcomes and the extent of relevant researcher ignorance created by missing data. For instance, missing peer x has different implications for study of the effect of peer mean x than for study of peer variance in x .

Prior information about the missingness process or the peer group formation process can strengthen inference by providing insight into the unobserved influence. In particular, this paper shows how to use knowledge that peer groups were formed by random assignment to point-identify and obtain unbiased estimates of certain types of peer effects without imposing conditions on the missingness process. Without any conditions on the sorting or missingness processes, peer effects are partially identified. With this insight, researchers can study the range of peer effects consistent with the model and observed data, which may be of interest in its own right. It also illuminates the identifying power of stronger conditions by contrast. One can go to the other extreme and assume enough structure on the missing data and unobserved influences to make likelihood and imputation methods available, but one would also introduce a greater risk of bias from model misspecification.

Many related methodological questions remain open. First, how can these identification

results about analysis of mean regression be generalized to cover other linear index models, such as models of binary outcomes or quantile regression? Second, how can effects of other properties of the peer distribution be analyzed? Peer effects are often small-group phenomena, where the particulars of the missing data matter and in which one cannot rely upon limits to reveal censored properties. Third, how can prior information about nonrandom sorting into peer groups be used to deal with missing peer data? Models that posit sorting on the sometimes-missing variable can leverage observed x for knowledge of missing peer x data. Unifying this with assumptions about sorting on influences that the econometrician never observes [Altonji et al. (2005), Rothstein (2009)] could prove useful. In some contexts, researchers may derive credible conditions on the missingness and sorting processes from models of individual behavior and properties of equilibria.

An important, related area for future attention to the challenge of missing data will be in the analysis of networks [Jackson (2007), Bramoullé et al. (2009), Christakis et al. (2010)]. In networks, each individual can be affected by all other individuals. Therefore, unless the complete network is fully observed, there is missing causally relevant data. This arises unavoidably given any sample of nodes within a network other than the full network and poses a real challenge for inference.

The methods developed here are applied to the STAR data to learn about peer effects in first grade classrooms with peer-influencing covariate data missing not at random. Without imposing any conditions on the missingness or sorting processes, there exist distributions of the missing data consistent with peer effect of mean lagged achievement ranging from -0.45 to an implausibly large 20.11. If there were no missing covariate data, a point estimate would be obtained without any additional assumptions.

This shows that missing peer data can create serious problems for inference because missing data do not remain isolated to those individuals. Bear this result in mind when interpreting IDP estimates of peer effects extant in the literature. When peer groups are formed by random assignment of individuals from given populations, IDP estimates attenuate to zero regardless of how the distribution of missing characteristics compares to the distribution of observed

ones. However, when groups are formed in other ways, the missingness process and group formation processes can interact in a wide range of ways that render IDP estimates biased and inconsistent with the magnitude and direction of the bias specific to the process.

We develop a set of peer-effect estimators to deal with missing data and compare it to approaches available in the literature. The unrestricted p -weight estimator is unbiased but imprecise and, in Monte Carlo simulations, performs even worse than the IDP when judged by mean square error. The restricted p -weight estimators proposed here dramatically improve precision at the expense of introducing bias. The bias appears minimal in simulations and these appear to outperform the unrestricted p -weight estimator and existing estimators — the IDP and AP-corrected IDP — on mean squared error.

Maintaining the assumption of random assignment of missing students to classes, we obtain point estimates suggesting substantial positive average effects of lagged peer achievement. Across all students, the unbiased estimates of the mean peer achievement effect including the richest set of conditioning covariates (including peer race, gender, and low-income status) yields a point estimate suggesting that raising mean peer lagged achievement by 10 percentile points would raise own average achievement by 2.5 percentile points, but the estimate is imprecise. Restricted versions of the estimator produce stable, significant estimates of average peer effects, suggesting that a 10 percentile point increase in peer achievement would increase own achievement by around 3.5 percentile points. Allowing for heterogeneous effects of the peer mean depending on one's own kindergarten achievement level, there is some evidence that students coming into first grade at lower levels of achievement benefit more than students coming in at higher levels of achievement from having higher-achieving peers.

If this were the proper specification of the production function and if agents' behavioral responses were stable, heterogeneous effects in this pattern would suggest that a policy with more achievement mixing would raise average achievement and reduce the difference in average outcomes between low, middle, and high kindergarten achievers compared to a policy with more achievement tracking. However, the analysis does not account for the effects of changes in the spread of the peer achievement distribution. Such effects may be important but are

more difficult to handle in the presence of missing data.

Although studying peer groups formed by random assignment helps deal with the selection problem and the missing-data problem, it limits understanding in at least two other ways. First, all observed variation in peer groups is derived only from small-sample variation in the process of random assignment to finite-sized groups. This creates tension with the desire to have like students assigned to a set of peer treatments embodying substantial variation. The problem worsens as group size increases, holding the number of groups fixed. Therefore, studies of peer effects in classrooms will suffer more than studies of peers in smaller groups, such as studies of college roommate effects [Sacerdote (2001), Zimmerman (2003)].

Second, the data are not directly informative about achievement tracking. Most observed classes represent small deviations from those that would be observed given a policy of intentional achievement mixing in classes. Figure 3 illustrates this fact by comparing the observed distribution of class means and standard deviations with those that are obtained when observed-data students are intentionally mixed or intentionally tracked within schools. Results obtained under random assignment — equivalent to small deviations from intentional mixing — may not be informative about what would occur under a policy of intentional tracking, which implies large deviations from mixing [Duflo et al. (2008), Carrell et al. (2010)]. Many pedagogical and social dynamics might change and affect the structure of peer effects. Project STAR did collect measures of teacher behavior in the classrooms but has not yet made them available. These measures could shed some light on pedagogical responses to differences in the peer distribution.

References

- Altonji, J., Elder, T., and Taber, C. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of Political Economy*, 113(1):151–184.
- Ammermueller, A. and Pischke, J.-S. (2008). Peer effects in european primary schools: Evidence from progress in international reading literacy study. LSE unpublished draft.
- Ammermueller, A. and Pischke, J.-S. (2009). Peer effects in european primary schools: Evidence from pirls. Forthcoming in *Journal of Labor Economics*.
- Arcidiacono, P., Foster, G., Goodpaster, N., and Kinsler, J. (2007). Estimating spillovers using panel data, with an application to the classroom. Working paper, Duke University.
- Arcidiacono, P. and Nicholson, S. (2005). Peer effects in medical school. *Journal of Public Economics*, 89(2-3):327–350.
- Atkinson, A., Burgess, S., Gregg, P., Propper, C., and Proud, S. (2008). The impact of classroom peer groups on pupil GCSE results. Centre for Market and Public Organization Working Paper 08/187.
- Betts, J. R. and Zau, A. (2004). Peer groups and academic achievement: Panel evidence from administrative data. Working paper, University of California – San Diego.
- Boggs, P., Byrd, R., and Schnabel, R. (1987). A stable and efficient algorithm for nonlinear orthogonal distance regression. *SIAM Journal on Scientific and Statistical Computing*, 8(6):1052–1078.
- Boozer, M. and Cacciola, S. E. (2001). Inside the “black box” of project star: Estimation of peer effects using experimental data. Yale Economic Growth Center Discussion Paper 832.
- Boucher, V., Bramoullé, Y., Djebbari, H., and Fortin, B. (2010). Do peers affect student achievement? evidence from canada using group size variation. CLSRN working paper, UBC Department of Economics.

- Boyd, D., Grossman, P., Lankford, H., Loeb, S., and Wyckoff, J. (2006). How changes in entry requirements alter the teacher workforce and affect student achievement. *Education Finance and Policy*, 1(2):176–216.
- Bramoullé, Y., Djebbari, H., and Fortin, B. (2009). Identification of peer effects through social networks. *Journal of Econometrics*, 150(1):41–55.
- Burke, M. A. and Sass, T. R. (2008). Classroom peer effects and student achievement. Federal Reserve Bank of Boston working paper 08-5.
- Carrell, S., Sacerdote, B., and West, J. (2010). Beware of economists bearing reduced forms? an experiment in how not to improve student outcomes. Working paper, University of California – Davis.
- Carroll, R. (2006). *Measurement error in nonlinear models: a modern perspective*. Chapman & Hall/CRC, Boca Raton, FL.
- Cascio, E. and Schanzenbach, D. W. (2008). Age and the education production function. NBER Working Paper 13663.
- Christakis, N., Fowler, J., Imbens, G., and Kalyanaraman, K. (2010). An Empirical Model for Strategic Network Formation. NBER Working Paper 16039.
- Cooley, J. (2006). Desegregation and the achievement gap: Do diverse peers help? Working paper, Duke University.
- Dee, T. S. (2004). Teachers, race, and student achievement in a randomized experiment. *Review of Economics and Statistics*, 86(1):195–210.
- Ding, W. and Lehrer, S. (2007). Do peers affect student achievement in china’s secondary schools? *Review of Economics and Statistics*, 89(2):300–312.
- Duflo, E., Dupas, P., and Kremer, M. (2008). Peer effects and the impact of tracking: Evidence from a randomized evaluation in Kenya. CEPR Discussion Paper 7043.

- Epple, D., Newlon, E., and Romano, R. (2002). Ability tracking, school competition, and the distribution of educational benefits. *Journal of Public Economics*, 83(1):1–48.
- Epple, D. and Romano, R. (1998). Competition between private and public schools, vouchers, and peer-group effects. *American Economic Review*, 88(1):33–62.
- Finn, J. D. and Achilles, C. M. (1999). Tennessee’s class size study: Findings, implications, misconceptions. *Educational Evaluation and Policy Analysis*, 21(2):97–109.
- Finn, J. D., Boyd-Zaharias, J., Fish, R. M., and Gerber, S. B. (2007). Health and Education Research Operative Services (HEROS), Inc.
- Finn, J. D., Gerber, S. B., Achilles, C. M., and Boyd-Zaharias, J. (2001). The enduring effects of small classes. *Teachers College Record*, 103(2).
- Foster, G. (2006). It’s not your peers, and it’s not your friends: Some progress toward understanding the educational peer effect mechanism. *Journal of Public Economics*, 90(8-9):1455–1475.
- Gardner, E. F. (1983). *Stanford Achievement Test Series: Multilevel Norms Booklet*. National Psychological Corporation, San Antonio, Texas.
- Gardner, E. F., Rudman, H. C., Karlsen, B., and Merwin, J. C. (1982). *Stanford Achievement Test Series*. Psychological Corporation, San Antonio, Texas.
- Graham, B. S. (2005). Identifying social interactions through excess variance contrasts. Working Paper, University of California, Berkeley.
- Graham, B. S. (2008). Identifying social interactions through conditional variance restrictions. *Econometrica*, 76(3):643–660.
- Guryan, J., Kroft, K., and Notowidigdo, M. (2009). Peer effects in the workplace: Evidence from random groupings in professional golf tournaments. *American economic journal. Applied economics*, 1(4):34–68.

- Hanushek, E. and Rivkin, S. (2006). *Teacher quality*, volume 2, pages 1051–1078.
- Hanushek, E. A. (1999). Some findings from an independent investigation of the tennessee star experiment and from other investigations of class size effects. *Educational Evaluation and Policy Analysis*, 21(2):143–163.
- Hanushek, E. A., Kain, J. F., Markman, J. M., and Rivkin, S. G. (2003). Does peer ability affect student achievement? *Journal of Applied Econometrics*, 18(5):527–544.
- Hidalgo-Hidalgo, M. (2007). On the optimal allocation of students when peer effect works: Tracking vs mixing. Working paper 07.14, Universidad Pablo de Olavide.
- Horowitz, J. L., Manski, C. F., Ponomareva, M., and Stöye, J. (2003). Computation of bounds on population parameters when the data are incomplete. *Reliable Computing*, 9(6):419–440.
- Hoxby, C. M. (2000). Peer effects in the classroom: Learning from gender and race variation. NBER Working Paper 7867.
- Hoxby, C. M. and Weingarth, G. (2007). Taking race out of the equation: School reassignment and the structure of peer effects. Working paper, Harvard University.
- Jackson, M. O. (2007). *The economics of social networks*, volume 1, pages 1–56. Cambridge University Press, Cambridge.
- Krueger, A. B. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics*, 114(2):497–532.
- Krueger, A. B. and Whitmore, D. M. (2001). The effect of attending a small class in the early grades on college-test taking and middle-school test results: Evidence from project STAR. *The Economic Journal*, 111(468):1–28.
- Lavy, V., Paserman, M. D., and Schlosser, A. (2007). Inside the black box of ability peer effects: Evidence from variation in high and low achievers in the classroom. NBER Working Paper W14415.

- Lee, L. (2007). Identification and estimation of econometric models with group interactions, contextual fixed factors and fixed effects. *Journal of Econometrics*, 140(2):333–374.
- Lefgren, L. (2004). Educational peer effects and the chicago public schools. *Journal of Urban Economics*, 56(2):169–191.
- Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies*, 60(3):531–542.
- Nye, B., Hedges, L. V., and Konstantopoulos, S. (2000). The effects of small classes on academic achievement: The results of the tennessee class size experiment. *American Educational Research Journal*, 37(1):123–151.
- Nye, B., Konstantopoulos, S., and Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, 26(3):237–257.
- Rothstein, J. (2009). Student sorting and bias in value-added estimation: Selection on observables and unobservables.
- Rothstein, J. (2010). Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement.
- Sacerdote, B. (2001). Peer effects with random assignment: Results for Dartmouth roommates. *The Quarterly Journal of Economics*, 116(2):681–704.
- Vigdor, J. and Nechyba, T. (2007). *Peer Effects in North Carolina Public Schools*. MIT Press, Cambridge, Mass.
- Whitmore, D. (2005). Resource and peer impacts on girls’ academic achievement: Evidence from a randomized experiment. *American Economic Review*, 95(2):199–203.
- Word, E. et al. (1990). The state of tennessee’s student/teacher achievement ratio (star) project: Technical report 1985-1990. Technical report, Tennessee State Department of Education.

- Zabel, J. E. (2008). The impact of peer effects on student outcomes in New York City public schools. *Education Finance and Policy*, 3(2):197–249.
- Zaharias, J. B., Achilles, C. M., and Cain, V. A. (1995). The effect of random class assignment on elementary students' reading and mathematics achievement. *Research in the Schools*, 2(2):7–14.
- Zimmerman, D. J. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. *Review of Economics and Statistics*, 85(1).

7 Tables

Table 1: **Comparison of estimators' performance in four different data-generating processes (DGP).** DGP vary along two dimensions: how the ordering of schools' missing student mean lagged achievement $E[v^m|s]$ was determined and whether teachers (correlated effects) are absent/present. For each DGP, 5,000 Monte Carlo replications were performed. In each replicate, all 7 estimators were used to estimate β . The distribution of each estimator's $\hat{\beta}$ s measure its squared error, variance, and mean squared error (MSE). Column 1 expresses each estimator's MSE as a multiple of the MSE for the unrestricted p -weight estimator in the no-teacher condition under the same $E[v^m|s]$ -ordering process. Columns 2 and 3 display the share of each estimator's MSE due to squared error and variance.

Order of schools' $E[v^m s]$:	Same as order of $E[v^{ob} s]$			Random		
	MSE(Estimator) MSE(Unrstr. p-wght.)	Sqrd Error MSE	Variance MSE	MSE(Estimator) MSE(Unrstr. p-wght.)	Sqrd Error MSE	Variance MSE
	1	2	3	1	2	3
Estimator:	No teachers					
Unrstr. p-weight, K=75	1.0000	0.0002	0.9998	1.0000	0.0003	0.9997
Restricted p-weight, K=25	0.5008	0.0006	0.9994	0.6903	0.0007	0.9993
Restricted p-weight, K=15	0.5015	0.0004	0.9996	0.6852	0.0005	0.9995
Restricted p-weight, K=3	0.4703	0.0002	0.9998	0.6584	0.0003	0.9997
Restricted p-weight, K=1	0.4284	0.0001	0.9999	0.6346	0.0001	0.9999
Indvdl.-deletion procd. (IDP)	0.5185	0.6786	0.3214	0.9463	0.5829	0.4171
AP-corrected IDP	0.6450	0.0011	0.9989	0.7773	0.0014	0.9986
With teachers whose quality variance is 0.4 times $V[v^{ob} s]$						
Unrstr. p-weight, K=75	1.5631	0.0002	0.9998	5.2817	0.0001	0.9999
Restricted p-weight, K=25	0.7294	0.0002	0.9998	3.3225	0.0001	0.9999
Restricted p-weight, K=15	0.7186	0.0001	0.9999	3.2457	0.0000	1.0000
Restricted p-weight, K=3	0.6669	0.0000	1.0000	3.0756	0.0000	1.0000
Restricted p-weight, K=1	0.5995	0.0000	1.0000	2.9432	0.0000	1.0000
IDP	0.5837	0.6081	0.3919	2.2333	0.2491	0.7509
AP-corrected IDP	0.8851	0.0005	0.9995	3.2982	0.0002	0.9998

Table 2: Variables and measures

y	Average of student's observed, nationally normed Stanford Achievement Test math, reading, and listening percentiles at the end of first grade (G1)
x	<i>Own covariates</i>
	<i>Own-demographic variables</i>
x_1	Same as y except from end of kindergarten (GK)
	Indicators that student is African-American, is female, and qualifies for free- or reduced-price lunch in G1.
	Both student's years of age on Sept. 30, 1986, and an indicator student was age-eligible for first-grade prior to the study year.
	<i>Class variables</i>
	Total number of students in each student's first grade class
	Indicator student in regular type class. Regular-with-aide type omitted.
	Teacher's years of experience
	Indicators that teacher has a master's or higher degree and that teacher is on a Tennessee career ladder step.
	<i>Peer demographic variables</i>
	Three variables measuring the fraction of each student's G1 classmates who are African-American, who are female, and who are low-income.
	<i>Other variables</i>
v	Variable by which student influences peers $\equiv g(x)$. Either x_1 or $(1(x_1 \leq 33.3), 1(x_1 \geq 66.7))'$.
z	Indicator that student's own v is observed
p	Fraction of student's G1 classmates with $z = 1$
1_s	Vector of school indicators

Table 3: Summary statistics for students in first grade, nonsmall classes

Variable	Mean	Std. Dev.	N or % observed	R ²	
				1 _s	1 _c
Individual variables					
G1 achievement percentile, y	50.7	24.7	96.9	0.28	0.33
GK achievement percentile, x_1	54.5	20.7	57.7	0.27	0.31
Indicator x_1 in bottom tercile	0.17		57.7	0.19	0.23
Indicator x_1 in top tercile	0.31		57.7	0.16	0.21
1(x_1 observed), z	0.58		4794	0.07	0.09
African-American indicator	0.33		99.4	0.70	0.71
Female indicator	0.48		99.7	0.01	0.02
Low-income indicator	0.52		97.1	0.29	0.31
Years of age	6.7	0.5	100	0.04	0.07
Old-for-grade indicator	0.21		100	0.04	0.08
First grade classroom peer variables					
Share of peers with x_1 observed, p	0.58	0.16	100	0.72	0.98
Mean x_1 among observed peers, \bar{x}_1^{ob}	54.4	11.9	100	0.87	0.99
Class-level variables					
Class size	23.1	2.4	100	0.82	1.00
Regular class indicator	0.53		100	0.08	1.00
Teacher has master's indicator	0.34		100	0.42	1.00
Teacher's years of experience	11.4	9.0	100	0.39	1.00
Teacher on career ladder indicator	0.71		100	0.46	1.00

Note: R^2 from regression of the variable on school dummies 1_s or class dummies 1_c.

Table 4: Effect of peer mean using unrestricted p -weight estimator among those with own-kindergarten (GK) achievement observed $z = 1$

Dependent variable: own first grade achievement y				
Independent variables	A	B	C	D
Observed peer mean GK achievement \times fraction of peers observed $\bar{x}_1^{ob} \cdot p$	-0.021 (0.213)	-0.010 (0.206)	0.042 (0.211)	0.121 (0.214)
Own GK achievement x_1	0.841*** (0.023)	0.794*** (0.024)	0.797*** (0.024)	0.800*** (0.024)
African-American indicator		-4.300*** (1.413)	-4.223*** (1.393)	-5.012*** (1.395)
Female indicator		1.115** (0.523)	1.123** (0.523)	-0.125 (0.679)
Low-income indicator		-5.004*** (0.819)	-5.078*** (0.819)	-4.694*** (0.831)
Regular-class indicator, with-aide omitted			-1.771 (1.301)	-0.516 (1.258)
p - school	Y	Y	Y	Y
School fixed effects	Y	Y	Y	Y
Class variables	N	N	Y	Y
Peer demographic variables	N	N	N	Y
Number of schools	75	75	75	75
Number of classes	210	210	210	210
Number of students	2684	2684	2684	2684
Adjusted R^2	0.36	0.69	0.69	0.69

Note: coefficient (within-class-corrected SE). Significance: *: 10% **: 5% ***: 1%

Specifications B-D also include years of age and old-for-grade indicator. Class and peer variables are listed in Table 2.

Table 5: Effect of mean peer lagged achievement \bar{v} using restricted p -weight with partitions of schools into K groups and AP-corrected estimators in the $z = 1$ sample

Dependent variable: own first grade achievement y				
Number of $p \times$ school-group interactions				
K	A	B	C	D
75: p -weight estimator	-0.021 (0.213)	-0.010 (0.206)	0.042 (0.211)	0.121 (0.214)
25: groups of 3 schools	0.148 (0.131)	0.189 (0.127)	0.239* (0.134)	0.231 (0.140)
15: groups of 5 schools	0.168 (0.147)	0.197 (0.145)	0.257* (0.148)	0.261* (0.153)
3: groups of 25 schools	0.186 (0.163)	0.207 (0.162)	0.241 (0.161)	0.253 (0.164)
1: all schools together	0.197 (0.160)	0.212 (0.159)	0.249 (0.155)	0.267* (0.158)
AP-corrected IDP	0.274 (0.189)	0.302 (0.189)	0.346* (0.183)	0.371** (0.184)
School fixed effects	Y	Y	Y	Y
Own lagged achievement	Y	Y	Y	Y
Own demographic variables	N	Y	Y	Y
Class variables	N	N	Y	Y
Peer demographic variables	N	N	N	Y
Number of schools	75	75	75	75
Number of classes	210	210	210	210
Number of students	2684	2684	2684	2684

Note: coefficient (within-class-corrected SE). Significance: *: 10% **: 5% ***: 1%

Each cell reports the estimate of β from a regression. Rows vary in the number of groups into which schools are partitioned for interaction with p . Columns vary in other covariates included. Specifications B-D also include years of age and an old-for-grade indicator.

Class and peer variables are listed in Table 2.

Table 6: Effect of peer mean using unrestricted p -weight estimator among all students

Dependent variable: own first grade achievement y				
Independent variables	A	B	C	D
Observed peer mean achievement \times fraction of peers observed $\bar{x}_1^{ob} \cdot p$	0.266 (0.205)	0.267 (0.201)	0.309 (0.199)	0.258 (0.205)
Own GK achievement $x_1 \times z$	0.850*** (0.023)	0.804*** (0.024)	0.806*** (0.024)	0.804*** (0.024)
African-American indicator		-9.497*** (2.120)	-9.499*** (2.098)	-10.211*** (2.187)
African-American indicator $\times z$		5.474** (2.635)	5.484** (2.586)	5.691** (2.602)
Female indicator		-0.084 (1.048)	-0.228 (1.054)	-0.473 (1.091)
Female indicator $\times z$		1.167 (1.146)	1.313 (1.150)	1.253 (1.150)
Low-income indicator		-9.085*** (1.288)	-8.934*** (1.295)	-8.856*** (1.319)
Low-income indicator $\times z$		4.140*** (1.379)	3.923*** (1.386)	3.867*** (1.385)
p : school	Y	Y	Y	Y
School fixed effects $\times z$	Y	Y	Y	Y
Other own-demographic $\times z$	N	Y	Y	Y
Class variables $\times z$	N	N	Y	Y
Peer demographic variables	N	N	N	Y
Number of schools	75	75	75	75
Number of classes	210	210	210	210
Number of students	4410	4410	4410	4410
Adjusted R^2	0.53	0.56	0.56	0.56

Note: coefficient (within-class-corrected SE). Significance: *: 10% **: 5% ***: 1%

Other own demographic, class, and peer variables are listed in Table 2.

Table 7: Effect of mean peer lagged achievement \bar{v} using estimators based on various partitions of schools into K -groups among all students

Dependent variable: own first grade achievement y				
Number of $p \times$ school-group interactions				
K	A	B	C	D
75: p -weight estimator	0.266 (0.205)	0.267 (0.201)	0.309 (0.199)	0.258 (0.205)
25: groups of 3 schools	0.281** (0.132)	0.321** (0.130)	0.374*** (0.125)	0.343*** (0.129)
15: groups of 5 schools	0.308** (0.150)	0.350** (0.150)	0.390*** (0.145)	0.378** (0.149)
3: groups of 25 schools	0.318* (0.163)	0.346** (0.165)	0.368** (0.159)	0.360** (0.164)
1: all schools together	0.313** (0.157)	0.339** (0.159)	0.365** (0.150)	0.362** (0.154)
AP-corrected IDP	0.403** (0.189)	0.427** (0.190)	0.450** (0.179)	0.450** (0.185)
School fixed effects $\times z$	Y	Y	Y	Y
Own lagged achievement $\times z$	Y	Y	Y	Y
Own demographic variables $\times z$	N	Y	Y	Y
Class variables $\times z$	N	N	Y	Y
Peer demographic variables $\times z$	N	N	N	Y
Number of schools	75	75	75	75
Number of classes	210	210	210	210
Number of students	4410	4410	4410	4410

Note: coefficient (within-class-corrected SE). Significance: *: 10% **: 5% ***: 1%

Each cell reports the estimate of β from a regression. Rows vary in the number of groups into which schools are partitioned for interaction with p . Columns vary in other covariates included. Specifications B-D also include years of age and an old-for-grade indicator.

Class and peer variables are listed in Table 2.

Table 8: Effect of peer mean assuming no teacher effects and known variance ratio

Dependent variable: own first grade achievement y				
Independent variables	A	B	C	D
Peer mean GK achievement	0.49*** (0.08)	0.56*** (0.08)	0.57*** (0.08)	0.60*** (0.08)
Own GK achievement	0.88*** (0.02)	0.84*** (0.02)	0.84*** (0.02)	0.83*** (0.02)
African-American indicator		-2.84** (1.25)	-3.01** (1.26)	-2.93** (1.26)
Female indicator		1.10** (0.52)	1.14** (0.52)	1.26 (0.58)
Low-income-family indicator		-5.00*** (0.66)	-5.01*** (0.66)	-4.89*** (0.66)
School fixed effects	Y	Y	Y	Y
Other own-demographic	N	Y	Y	Y
Class variables	N	N	Y	Y
Peer demographic variables	N	N	N	Y
Number of schools	75	75	75	75
Number of classes	210	210	210	210
Number of students	2684	2684	2684	2684

Note: coefficient (SE). Significance: *: 10% **: 5% ***: 1%

Estimated by GMM. Other own demographic, class, and peer variables listed in Table 2.

Table 9: Heterogeneous effects of peer mean among those with own-kindergarten achievement observed with unrestricted p -weight estimator

Dependent variable: own first grade achievement y				
Independent variables	Peer effect $\hat{\beta}_T$ if own GK achievement in tercile T			
	A	B	C	D
Own-GK achv. in bottom tercile	-0.03 (0.25)	0.02 (0.25)	0.22 (0.26)	0.34 (0.28)
Own-GK achv. in middle tercile	0.04 (0.24)	0.08 (0.23)	0.19 (0.25)	0.32 (0.27)
Own-GK achv. in top tercile	-0.11 (0.24)	-0.04 (0.24)	0.10 (0.25)	0.24 (0.28)
p : school	Y	Y	Y	Y
Tercile \times school fixed effects	Y	Y	Y	Y
Tercile \times own-GK achievement	Y	Y	Y	Y
Tercile \times other own variables	N	Y	Y	Y
Tercile \times class variables	N	N	Y	Y
Tercile \times peer demographic variables	N	N	N	Y
Number of schools	75	75	75	75
Number of classes	210	210	210	210
Number of students	2684	2684	2684	2684

Note: coefficient (within-class-corrected SE). Significance: *: 10% **: 5% ***: 1%

Full interaction of covariates with own-tercile. For terciles 2 and 3, estimates are main peer effect plus effect of peer-by-own-tercile indicator.

Table 10: Heterogeneous effects of peer mean by own-kindergarten-achievement tercile using different estimators

Dependent variable: own first grade achievement y			
Estimator	$\hat{\beta}_T$ for students with own kindergarten achievement in tercile:		
	Bottom	Middle	Top
Unrstr. p -weight, $K = 75$	0.337 (0.283)	0.316 (0.271)	0.239 (0.277)
Restricted p -weight, $K = 25$	0.439** (0.188)	0.323** (0.163)	0.213 (0.180)
Restricted p -weight, $K = 15$	0.505*** (0.187)	0.334** (0.168)	0.286 (0.184)
Restricted p -weight, $K = 3$	0.520*** (0.193)	0.322 (0.171)	0.334 (0.187)
Restricted p -weight, $K = 1$	0.547*** (0.198)	0.358** (0.162)	0.345** (0.170)
AP-corrected IDP	0.839*** (0.293)	0.550** (0.230)	0.045 (0.2710)

Note: coefficient (within-class-corrected SE). Significance: *: 10% **: 5% ***: 1%

Full interaction of covariates with own-tercile. For terciles 2 and 3, estimates are main peer effect plus effect of peer-by-own-tercile indicator.

Each row reports peer effect estimates by own-tercile with listed estimator from specification D as in Table 9.

8 Figures

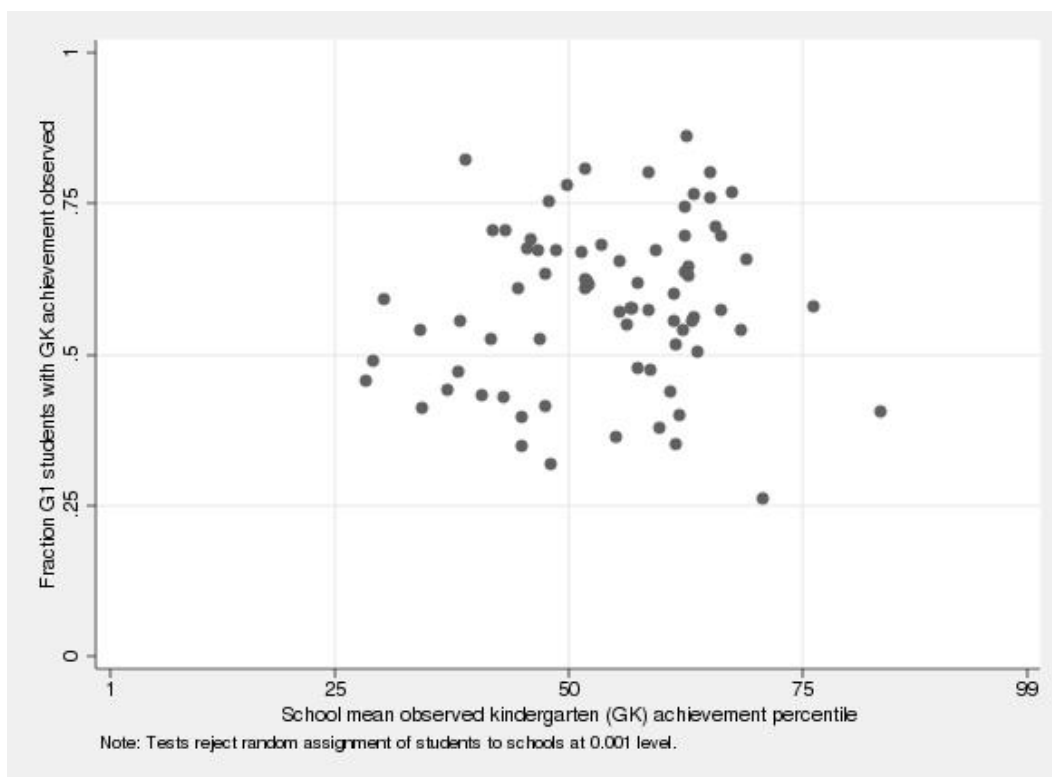


Figure 1: Students self-selected into schools. Restrict attention to students in first grade (G1), nonsmall classes. This figure presents the joint distribution of two variables across schools. The horizontal axis describes the mean observed GK achievement percentile among each school's first graders. The vertical axis describes each school's fraction of first grade students with kindergarten (GK) achievement percentile observed. F -tests reject the null hypothesis that students were randomly assigned to schools.

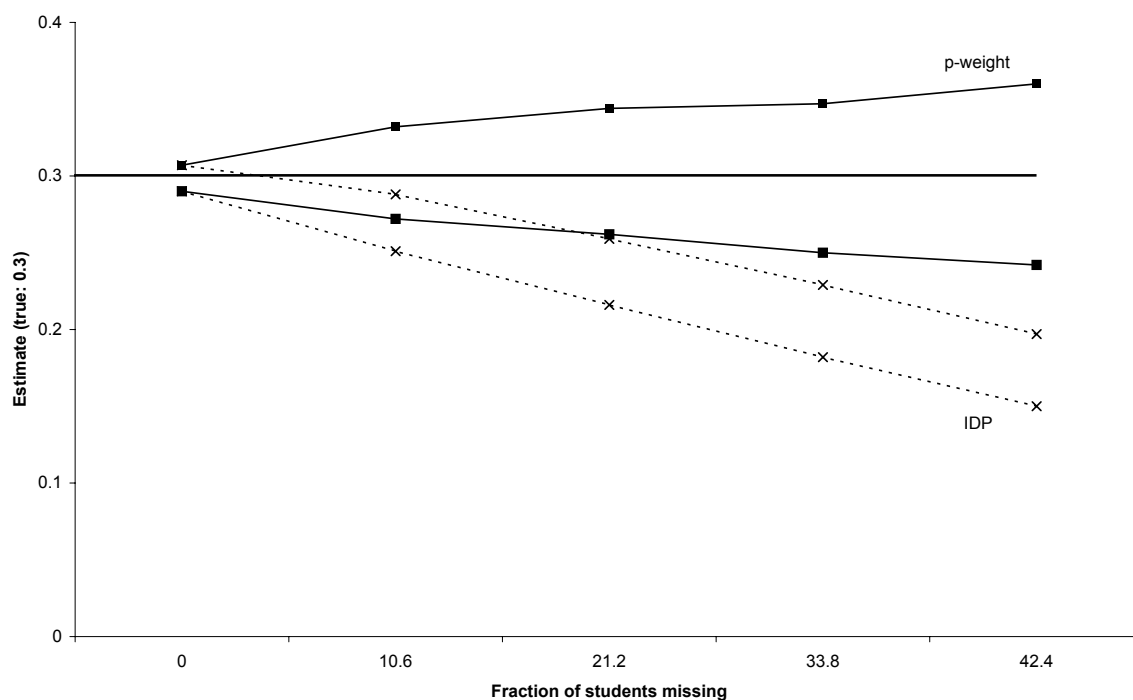


Figure 2: Comparison of IDP and p -weight estimators' performance in simulations with various fractions of lagged achievement data missing. The graph displays 95 percent confidence intervals for each estimator from 200 replications at each level of missingness. The IDP estimator attenuates to zero as the fraction of data increases. The p -weight estimator's interval always includes the true value of 0.3.

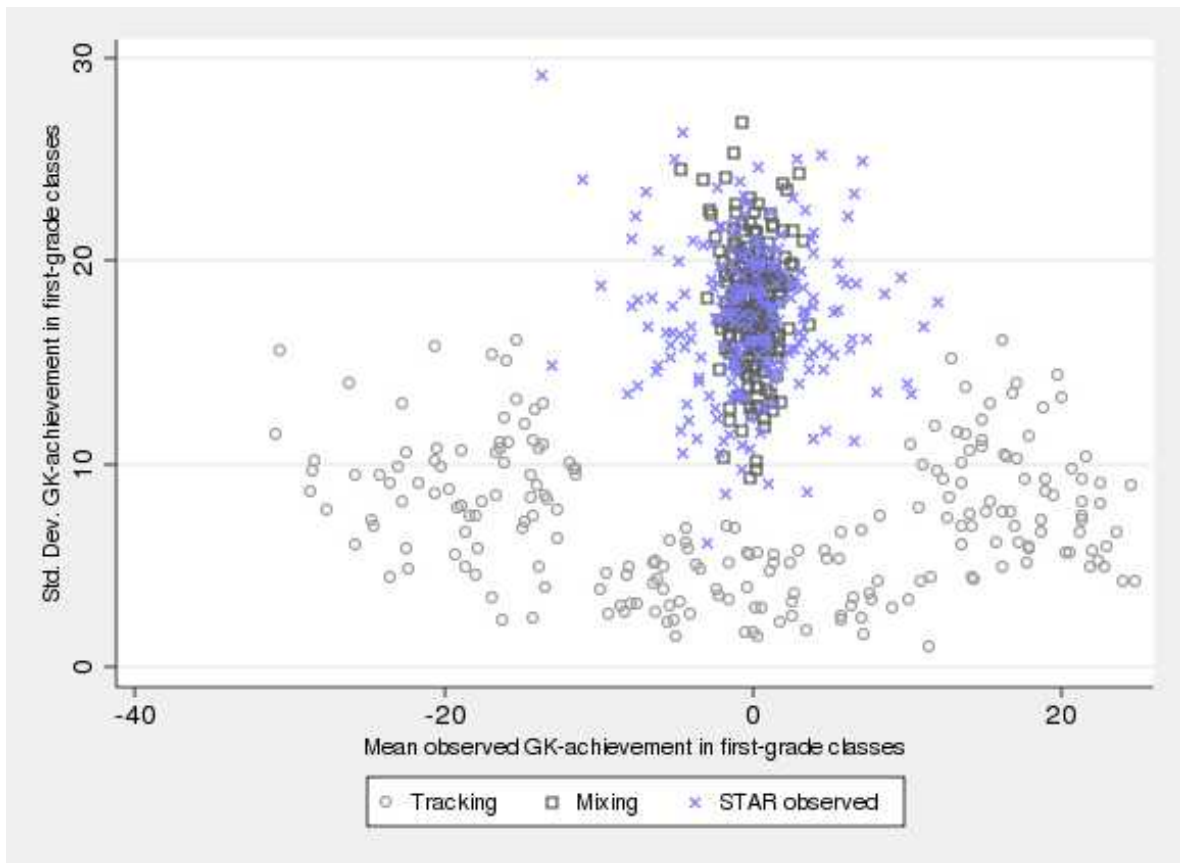


Figure 3: Comparison of classroom mean and standard deviation observed GK achievement under three alternative assignment policies: intentional mixing, intentional tracking, and that observed in STAR data (a realization of random assignment). Each point represents a classroom. Measures are centered within school. Mixing and tracking were accomplished by rank ordering observed students within school and assigning to the observed number of classes according to the appropriate policy.